*Animal Learning & Behavior*
*2001, 29 (4), 367-380*

# Recovery from the overexpectation effect: Contrasting performance-focused and acquisition-focused models of retrospective revaluation

AARON P. BLAISDELL
*University of California, Los Angeles, California*

JAMES C. DENNISTON
*Appalachian State University, Boone, North Carolina*

and

RALPH R. MILLER
*State University of New York, Binghamton, New York*

In four Pavlovian conditioned lick-suppression experiments, rats had two conditioned stimuli (CSs X and A) independently paired with footshock, followed by pairings of a compound of A and X with the footshock. On subsequent tests with CS X, less conditioned suppression was observed than in control subjects that lacked the compound AX→footshock trials. This *overexpectation* effect was reversed through posttraining extinction of CS A, a result consistent with both performance- and acquisition-focused models of retrospective revaluation. However, only performance-focused models could account for how posttraining increases or decreases in the A–footshock temporal interval attenuate the overexpectation effect.

The *overexpectation* effect in Pavlovian conditioning is produced by pairing a compound of two conditioned stimuli (CSs) A and X with an unconditioned stimulus (US) following independent pairings of each CS alone with the US. The typical result is reduced responding to the target CS (X) following compound conditioning, relative to a control group that did not receive AX→US pairings (Kamin & Gaioni, 1974; Khallad & Moore, 1996; Kremer, 1978; Levitan, 1975; Rescorla, 1970; but see St. Claire-Smith & Mackintosh, 1974). The overexpectation effect was originally touted as a successful prediction of the Rescorla–Wagner (1972) model of associative learning. The Rescorla–Wagner model attributes the overexpectation effect to a discrepancy between the expected and the observed USs during compound conditioning. In this framework, during initial training both CSs (A and X) acquire near-asymptotic levels of associative strength through separate pairings of each cue with the US. The subsequent compound trials produce an overexpectation of the US; that is, the sum of the separate associative strengths of A and X is larger by about a factor of two than the US experienced on each compound trial. Thus, during compound conditioning, both CSs are expected to lose associative strength until their sum equals the US experienced by the subject. This decrement in associative strength should result in a loss of conditioned responding.

In contrast with the acquisition-focused approach of Rescorla and Wagner (1972), performance-focused models attribute the overexpectation effect to a failure to express an acquired target CS–US association. One such performance-focused model is the comparator hypothesis, which is a response rule for the expression of Pavlovian associations (Denniston, Savastano, & Miller, 2001; Miller & Matzel, 1988). According to the comparator hypothesis, responding to a target CS (i.e., the CS presented at test) is determined by how well it predicts the US *relative to* the predictive value of other stimuli that were present during training with the target CS. Such a stimulus is hereafter called the target CS's comparator stimulus and may be the training context or other salient, discrete stimuli. The effectiveness of the comparator stimulus in attenuating responding to the target CS is determined not only by its predictive value, but also by the similarity of the US that it predicts to the US predicted by the target CS, with *similarity* including intrinsic qualities (e.g., US intensity and duration) and extrinsic qualities, such as spatiotemporal location relative to other stimuli in the situation (e.g., CS–US temporal interval). Specifically, three functional associations are presumed to determine conditioned responding in the presence of the target CS: (1) a target CS–US association, (2) a target-CS–comparator-stimulus association, and

Support for this research was provided by NIMH Grant 33881. We thank Francisco Arcediano, Daniel Burger, Raymond Chang, Martha Escobar, and Hernán Savastano for comments on a preliminary draft of this manuscript. Thanks are also due James Esposito and Gibbi George for assistance with the collection of data. Correspondence concerning this research should be addressed to R. R. Miller, Department of Psychology, SUNY, Binghamton, NY 13902-6000 (e-mail: rmiller@binghamton.edu).

*Note—Vincent LoLordo served as the action editor for this manuscript.*

 

(3) a comparator-stimulus–US association (illustrated in Figure 1). Presumably, these associations allow the presentation of one associate to activate not only a representation of the other associate, but also the temporal (and spatial) relationship between the two associates (Barnet, Grahame, & Miller, 1993; Savastano & Miller, 1998). Presentation of the target CS at test directly activates a spatiotemporally specific representation of the US by means of the target-CS–US association (Link 1). The target CS also activates a spatiotemporally specific representation of the US through an indirect pathway consisting of the target-CS–comparator-stimulus association (Link 2) and the comparator-stimulus–US association (Link 3), with the temporal location of the indirectly activated US representation relative to the CS being a summative function of the temporal relationships encoded within Links 2 and 3. The strengths, temporal locations, and other qualities of the US representations activated through the direct (Link 1) and indirect (Links 2 and 3) pathways are compared at test, and the outcome of this comparison determines the nature and magnitude of the conditioned response. Excitatory control of responding by a CS is presumed to decrease (and inhibitory control increase) as the strength and spatiotemporal similarity of the indirectly activated US representation increases relative to the directly activated US representation. Thus, conditioned responding to the target CS is attenuated at test by the associative status of its comparator stimulus (with respect to both the CS and the US—i.e., Links 2 and 3) and the similarity between the temporal relationship of the CS to the indirectly activated US representation and the temporal relationship of the CS to the directly activated US representation.

The comparator hypothesis accounts for the overexpectation effect as follows: As a result of the many compound AX→US pairings, CS A comes to act as the primary comparator stimulus for the target CS X (i.e., there is a strong Link 2 for A as X's comparator stimulus). The product of the strong X–A (Link 2) and strong A–US (Link 3) associations is large, thereby indirectly activating a strong US representation, which attenuates excitatory conditioned responding to CS X. In the control group, for which the compound of A and X is not paired with the US during Phase 2, the training context continues to serve as the primary comparator stimulus for X. Because the training context is less excitatory than CS A is (the latter being more salient and more highly correlated with the US than is the context), the context should be less effective in attenuating responding to CS X. Thus, the overexpectation group should exhibit less conditioned responding than the control group.

Although both the Rescorla–Wagner (1972) model and the comparator hypothesis can account for overexpectation, these two models make contrasting predictions regarding experimental manipulations conducted after overexpectation treatment (hereafter called posttraining manipulations). First, the comparator hypothesis predicts that posttraining extinction of the competing stimulus (A) should enhance conditioned responding to the target CS (X), whereas the Rescorla–Wagner model does not. In the framework of the comparator hypothesis, extinction of A should reduce the strength of the A–US association (Link 3). Demonstrations of recovery from related cue-competition phenomena, such as overshadowing (Harris & Westbrook, 1998; Kaufman & Bolles, 1981; Matzel, Schachtman, & Miller, 1985), blocking (Blaisdell, Gunther, & Miller, 1999), and the relative stimulus validity effect (Cole, Barnet, & Miller, 1995a) through extinction of the comparator stimuli in each case have offered support for the comparator hypothesis over the Rescorla–Wagner model interpreta-

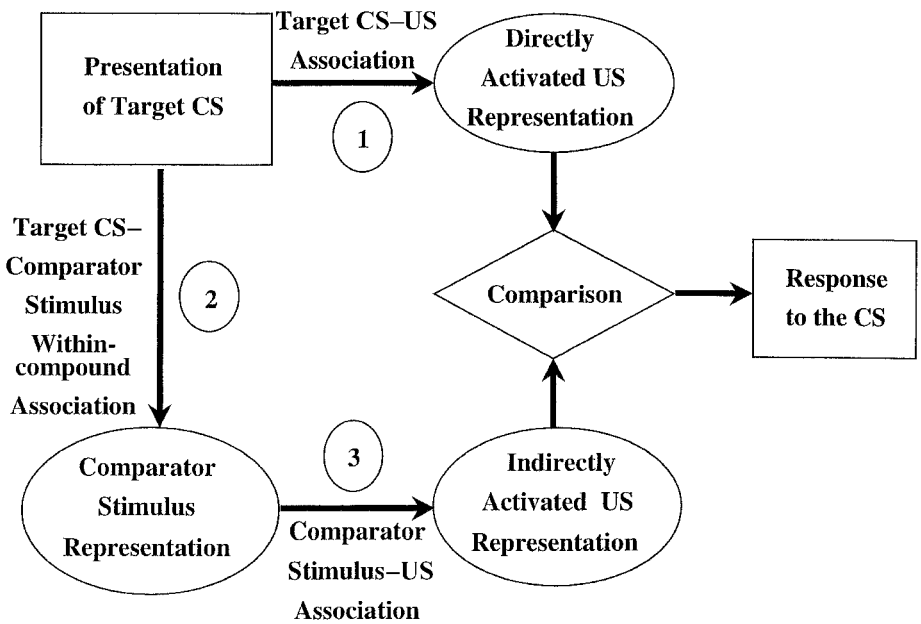


**Figure 1. The comparator hypothesis. See the text for details.**

tion of cue-competition effects (but some failures to observe such effects have been reported; e.g., Holland, 1999; Rauhut, McPhee, & Ayres, 1999; Rauhut, McPhee, DiPietro, & Ayres, 2000; Williams, 1996). We postpone discussion of more recently formulated acquisition-focused models of retrospective revaluation effects until the General Discussion section.

A second point of contrast for the two models involves the effects of posttraining modifications of associations to the comparator stimulus. For example, Blaisdell, Denniston, and Miller (1999) have shown that, *after* completion of overshadowing training (in which a more salient CS A overshadows a less salient CS X), shifting the A–US temporal relationship (e.g., from trace to delay or from delay to trace) attenuates the overshadowing deficit. As was stated above, the effectiveness of the comparator process is a direct function of the similarity between the directly and the indirectly activated US representations. Any manipulation that decreases the similarity of these US representations (e.g., a temporal shift in the comparator-stimulus–US association) should attenuate the modulatory effects of the comparator stimulus on responding to the target CS. Thus, posttraining qualitative (e.g., temporal) changes in the comparator-stimulus–US association should ameliorate cue-competition effects, such as overshadowing and overexpectation. Notably, contemporary acquisition-focused interpretations of cue competition fail to predict any effect of posttraining qualitative changes to the competing stimulus. If a target-CS–US association was not acquired during training, there should be nothing for posttraining manipulations to reveal. Therefore, reversal of overexpectation through posttraining shifts in the A–US temporal relationship after both phases of overexpectation treatment would support the (performance-focused) comparator hypothesis, but not the existing acquisition-focused models of acquired behavior. Although our laboratory has previously reported studies demonstrating the effects on responding to a CS of posttraining modifications in strength or content of the association between the CS's comparator stimulus and the US (e.g., Burger, Mallemat, & Miller, 2000), the present experiments speak to the generality of the underlying comparator process and are novel in illuminating the nature of the overexpectation effect.

Experiment 1 served as a demonstration of the overexpectation effect, using a conditioned lick-suppression procedure with rats. The purpose of Experiment 2 was to determine whether the overexpectation effect could be attenuated through posttraining extinction of the competing (A) stimulus. Finally, in Experiments 3 and 4, the effects of posttraining shifts in the A–US temporal relationship on conditioned responding to X were investigated.

## EXPERIMENT 1

Experiment 1 was designed to demonstrate the overexpectation effect. Subjects were randomly assigned to one of four groups (see Table 1). In Phase 1, all the subjects received, separately, 6 A→US pairings, 6 X→US pairings, and 6 nonreinforced presentations of CS B, with the three trial types interspersed during each session. In Phase 2, the subjects were administered the following treatments: (1) Group Overexpectation (OX) received 18 pairings of the AX compound with the US, (2) Group Con1 received 18 pairings of the BX compound with the US, (3) Group Con2 received 18 X→US pairings, and (4) Group Con3 received 18 B→US pairings. According to both the Rescorla–Wagner (1972) model and the comparator hypothesis, Group OX should show little responding to CS X relative to the three control groups, thereby demonstrating overexpectation. Group Con1 controlled for a generalization decrement interpretation of overexpectation by training CS X in compound with a stimulus (B) that had not previously signaled the US and testing on CS X alone. If overexpectation was merely a product of a failure to generalize responding from compound training in Phase 2 to testing on X alone, Group Con1 should also show a decrement in responding to X. Group Con2 served as an acquisition control that received an equivalent number of X→US pairings as Group OX. Group Con2 also controlled for the effects of US habituation as a possible source of low responding in Group OX. Finally, Group Con3 served to assess the behavioral control gained by CS X during the six Phase 1 X→US pairings.

### Method

#### Subjects

Forty-eight male (285–450 g) and 48 female (225–310 g) naive, Sprague-Dawley–descended rats, bred in our colony from Holtzman stock (Madison, WI), served as subjects. The subjects were individually housed in wire-mesh cages in a vivarium maintained on a 16:8-h light:dark cycle. Manipulations were performed approximately midway through the light portion of the cycle. A progressive water deprivation schedule was imposed over the week prior to the beginning of the experiment, until water availability was limited to 20 min per day. All the animals were handled three times per week for 30 sec, from time of weaning to the initiation of the study. The subjects were randomly assigned to one of four groups, ($n$s = 24), counterbalanced for sex.

#### Apparatus

Two types of experimental chambers were used. Chamber R was rectangular in shape and measured 24 × 9 × 16 cm (l × w × h). The walls and ceiling of the chamber were constructed of clear Plexiglas, and the floors were constructed of stainless steel rods measuring 0.5 cm in diameter, spaced 1.5 cm center-to-center, connected through NE-2 neon bulbs, which allowed for the delivery of constant-

**Table 1**
**Design Summary: Experiment 1**

| Group | Phase 1 | Phase 2 | Test |
|---|---|---|---|
| OX | A→US/X→US/B– | AX→US | X |
| Con1 | A→US/X→US/B– | BX→US | X |
| Con2 | A→US/X→US/B– | X→US | X |
| Con3 | A→US/X→US/B– | B→US | X |

Note—OX refers to overexpectation treatment. Con1, Con2, and Con3 refer to control Groups 1, 2, and 3, respectively. A, X, and B denote the noise, tone, and click stimuli, counterbalanced within groups. US represents the footshock unconditioned stimulus, and the slash (/) separates events that were interspersed.

current footshock produced by a high-voltage AC circuit in series with a 1.0-MΩ resistor. Each of six copies of Chamber R were housed in separate sound- and light-attenuating environmental isolation chests. Enclosure R was dimly illuminated by a 2-W (nominal at 120 VAC) incandescent houselight driven at 60 VAC and mounted on the back wall of the conditioning chamber.

Chamber V was a 25.5-cm-long box in the shape of a vertical bottom-truncated V. The chamber was 28 cm high, 21 cm wide at the top, and narrowed to 5.25 cm wide at the bottom. The ceiling was constructed of clear Plexiglas, the front and back walls were black Plexiglas, and the side walls were stainless steel. The floor consisted of two 25.5-cm-long parallel metal plates, each 2 cm wide, separated by a 1.25-cm gap. A constant-current footshock could be delivered through the metal walls and floor of the chamber. Each of six copies of Chamber V was housed in separate sound- and light-attenuating environmental isolation chests. Chamber V was illuminated by a 7-W (nominal at 120 VAC) bulb driven at 60 VAC. This bulb was mounted on the inside wall of the environmental enclosure, approximately 30 cm from the center of the experimental chamber, with the light entering the chamber primarily by reflection from the ceiling of the environmental enclosure. The light intensities in the two chambers were approximately equal, owing to differences in the opaqueness of the walls in Chambers R and V. Each subject received all of its treatment in a single chamber type (i.e., R or V), unless specifically stated in the text. Chambers were counterbalanced within groups so that, in each group, half of the animals received their treatment in Chamber R and the rest of the animals received their treatment in Chamber V.

Chambers R and V were both equipped with a water-filled lick tube which extended about 1 cm into a cylindrical niche (4.5 cm in diameter) from the back of the niche. The niche was located at one end of the chamber, with its bottom approximately 2 cm above the floor. Three 45-Ω speakers on three different inside walls of the environmental enclosures could deliver either a high-frequency complex tone (T; 3000 and 3200 Hz) 8 dB (A-scale) above background, a white noise (N) stimulus 8 dB (A-scale) above background, or a 6/sec click train (C) stimulus 8 dB (A-scale) above background. The noise, click, and tone, served as CSs A, X, and B, counterbalanced within groups. A 1.0-mA footshock could be delivered through the chamber floors. All CSs were 10 sec in duration. The footshock US was 0.5 sec in duration. For all the shocked trials, US onset occurred simultaneously with CS termination. Ventilation fans in each enclosure provided a constant 76-dB (A-scale) background noise.

### Procedure

Table 1 summarizes the critical aspects of the treatment procedure. All the sessions were 1 h in duration, except where otherwise noted below. Experiment 1 was performed in two replications, counterbalanced within groups. Because no main effect or interaction with replication was obtained, no further mention of replication is made.

**Acclimation.** On Day 1, the subjects were acclimated to the conditioning apparatus. In order to establish a stable baseline rate of drinking, lick tubes were in place, providing the subjects with water throughout the session; other than the lick tubes, the subjects received no nominal stimulus presentations.

**Phase 1 (element training).** Prior to Phase 1, lick tubes were removed from the apparatus. On Days 2 and 3, all the subjects received three A→US trials, three X→US trials, and three B− trials per daily session, with an average intertrial interval of 6 ± 3 min, for a total of six trials with each stimulus.

**Phase 2 (compound training).** On Days 4–6, the subjects in Group OX received 6 daily AX→US trials (A and X had common onset and termination), Group Con1 received 6 daily BX→US trials (B and X had common onset and termination), Group Con2 received 6 daily X→US trials, and Group Con3 received 6 daily B→US trials, for a total of 18 trials per group. All the trials were presented with a mean intertrial interval of 11 ± 4 min.

**Reacclimation.** On Days 7 and 8, lick tubes were reinstalled, and the subjects were allowed to drink during each daily 60-min session to reestablish baseline levels of drinking. These sessions did not have any nominal stimulus presentations.

**Testing.** On Day 9, the subjects were tested for conditioned lick suppression to CS X by presenting the stimulus immediately after completion of the first 5 cumulative seconds of licking. Thus, all the subjects were licking at the time of CS onset. Data from those subjects failing to complete the first 5 cumulative seconds of licking within 60 sec (which demonstrates an unusually great fear of the context) were eliminated from the study. One subject from Group Con2 met this criterion of elimination. In addition, owing to experimenter error, 12 subjects (3 from each group) received incorrect treatment during Phase 2; thus, their data were excluded from all statistical analyses. Test sessions were 15 min maximum in duration.

## Results and Discussion

Experiment 1 successfully demonstrated overexpectation—that is, weaker conditioned responding to X was observed in Group OX than in each of the control groups.

A one-way analysis of variance (ANOVA) conducted on pre-CS times to complete 5 cumulative seconds of licking in the absence of X revealed no group differences [$F(3,79) = 1.05$, $MS_e = 0.014$, $p > .35$]. Thus, groups did not differ in their fear to the context going into the test with the CS. A one-way ANOVA conducted on time to complete 5 cumulative seconds of licking in the presence of X found an effect of treatment group [$F(3,79) = 3.60$, $MS_e = 0.310$, $p < .02$]. Planned comparisons were conducted using the error term from the analysis of suppression to CS X. Although Group OX suppressed less to X than did Groups Con1 and Con2 [$F$s$(1,79) > 5.24$, $p$s $< .03$], Groups OX and Con3 did not differ significantly in suppression to X [$F(1,79) = 3.21$, $p > .08$].

The failure to detect a difference between these two last groups was likely due to the unusually high within-groups variance ($MS_e = 0.310$). Analysis of lick suppression in our laboratory sometimes finds that pre-CS scores and CS scores are correlated. Although we attempted to extinguish generalized fear to the test context, residual anxiety (i.e., diffuse fear) was likely present. The tendency not to drink produced by this anxiety may have summated with associative CS-induced conditioned suppression of drinking. We used an analysis of covariance (ANCOVA), with the intent of factoring out the stimulus-nonspecific fear. All ANCOVAs were conducted on log latencies to complete 5 cumulative seconds of licking in the presence of X, using log pre-CS scores as a covariate. (For consistency, an ANCOVA was used for all statistical tests of suppression of licking to X in subsequent experiments.) This analysis revealed an effect of treatment group [$F(3,78) = 4.65$, $MS_e = 0.237$, $p < .005$; see Figure 2]. Planned comparisons were conducted using the error term from the ANCOVA. These comparisons revealed lower suppression in Group OX than in Group Con1 [$F(1,78) = 7.42$, $p < .01$], Group Con2 [$F(1,78) = 11.72$, $p < .001$], and Group Con3 [$F(1,78) = 7.57$, $p < .01$]. The relatively

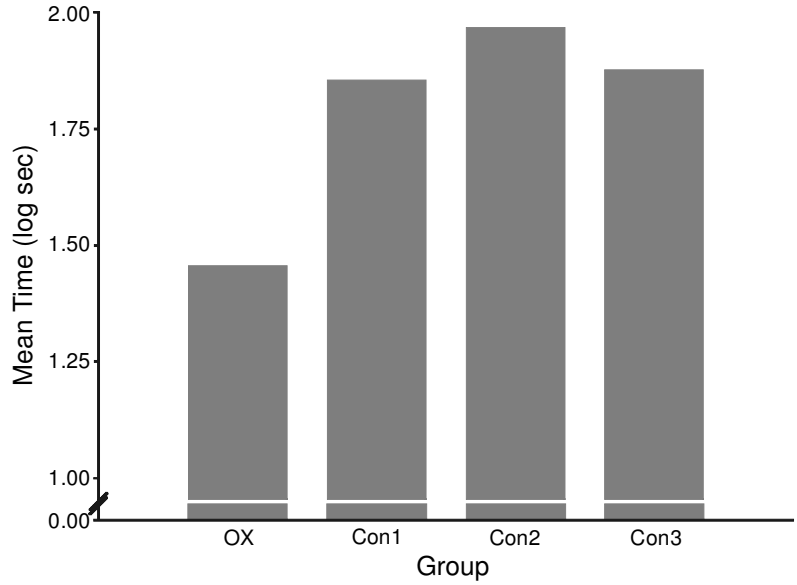


**Figure 2. Experiment 1: Adjusted mean times (log sec) to complete 5 cumulative seconds of licking in the presence of the target CS (X). For the procedure, see Table 1.**

strong suppression to X by Group Con1 ruled out a generalization decrement account of the low suppression in Group OX. Strong suppression by Group Con2 indicated that habituation to the US did not contribute significantly to the overexpectation effect (as is demonstrated here). Finally, strong suppression by Group Con3 demonstrated that the six X→US pairings of Phase 1 were sufficient to support strong conditioning to X.

Thus, Experiment 1 demonstrated the overexpectation effect with rats, using a conditioned lick-suppression procedure. That is, relative to control groups lacking in one or another aspect of overexpectation treatment, rats receiving separate A→US and X→US pairings followed by compound AX→US pairings exhibited reduced suppression to X. Inspection of response levels in the three control groups in Figure 2 suggests that X was near a performance ceiling at the end of Phase 1. Responding in Group Con2, which received 18 X–US pairings in Phase 2 after receiving 6 X–US pairings in Phase 1, was not much higher than responding in Group Con3, which received only the Phase 1 treatment. Presumably, if given few enough Phase 1 pairings, responding to X would be lower in Group Con3 than in Group Con2.

## EXPERIMENT 2

The purpose of Experiment 2 was to determine whether the overexpectation deficit could be reversed by massively extinguishing stimulus A (presumably X's comparator stimulus), as has been observed with other forms of cue competition. If, as is predicted by the comparator hypothesis, the attenuation of responding to X after AX→US pairings is due to competition with a strong A–US association (Link 3) at the time of testing, extinction of the A–US association should recover strong responding to X. This same logic has prevailed in elucidating the important role served by comparator stimuli in modulating responding to the target CS in many cue-competition preparations (e.g., Blaisdell, Gunther, & Miller, 1999; Cole et al., 1995a; Kaufman & Bolles, 1981; Matzel et al., 1985). Such demonstrations have previously rallied support for performance-focused models of cue-competition effects, such as the comparator hypothesis (e.g., Miller & Matzel, 1988; Miller & Schachtman, 1985), over acquisition-focused models of acquired behavior. As was previously mentioned, recent revisions of some acquisition-focused models—specifically, Van Hamme and Wasserman's (1994) revision of the Rescorla–Wagner (1972) model and Dickinson and Burke's (1996) revision of Wagner's (1981) SOP model—also anticipate extinction-mediated recovery from cue competition. But discussion of these two models will be postponed until the General Discussion section.

In the present experiment, all the rats received separate A→US, X→US, and B→US pairings in Phase 1. In Phase 2, the rats in Groups OX.A and OX.B received AX→US treatment (i.e., overexpectation training), whereas the rats in Group Con.B received CX→US pairings (C was an associatively neutral CS prior to Phase 2 treatment and thus served as a control for generalization decrement). Finally, in Phase 3, the rats in Group OX.A received extinction of A, whereas the rats in Groups OX.B and Con.B received equivalent extinction of B. If the overexpectation effect (i.e., weak responding to X in Group OX.B, relative to Group Con.B) is due to the presence of a strong A–US as-

sociation (Link 3) in conjunction with a strong X–A association, then on the basis of the comparator hypothesis (as well as other models to be discussed later), extinction of A (Group OX.A) should attenuate the overexpectation deficit in Group OX.A—that is, induce strong conditioned responding to X.

## Method

### Subjects

Thirty-six male (265–430 g) and 36 female (200–350 g) naive, Sprague-Dawley–descended rats, bred in our colony from Holtzman stock, served as subjects. The animals were randomly assigned to one of three groups (*n*s = 24), Groups OX.A, OX.B, and Con.B, counterbalanced for sex. The animals were housed and maintained as in Experiment 1. The apparatus and stimuli were identical to those used in Experiment 1, except for the addition of a buzzer that could be presented at 8 dB (A-scale) above background. The noise, tone, and buzzer served as CSs A, B, and C, counterbalanced within each group. The clicks served as CS X for all the subjects.

### Procedure

Table 2 summarizes the critical aspects of the procedure, along with an explanation of the group names. Like Experiment 1, this experiment was performed in two balanced replications. No significant main effect or interaction with replication was obtained.

**Acclimation.** On Day 1, the subjects were acclimated to the context as in Experiment 1.

**Phase 1 (element training).** Prior to Phase 1, lick tubes were removed from the apparatus. On Days 2 and 3, all the subjects received three A→US trials, three X→US trials, and 3 B→US trials per daily 75-min session, for a total of six trials with each CS. Trials were pseudorandomly distributed in the session with an interval of 8 ± 3 min.

**Phase 2 (compound training).** On Days 4–6, the subjects in Group OX.A and OX.B received 6 daily AX→US trials, and the subjects in Group Con.B received 6 daily CX→US trials, both with a mean intertrial interval of 11 ± 4 min, for a total of 18 trials per group across the daily 60-min sessions.

**Phase 3 (extinction training).** On Days 7–10, the subjects in Group OX.A received 50 daily A− (i.e., extinction) trials, whereas the subjects in Group OX.B and Con.B received an equivalent number of B− trials, for a total of 200 trials. Trials were pseudorandomly distributed in the daily 60-min session, with an intertrial interval of 1 min ± 30 sec. We did not assess responding to CS A because we had previously demonstrated, under similar training, extinction, and test conditions, that 200 extinction trials were more than sufficient to extinguish responding to A (Blaisdell, Gunther, & Miller, 1999). Furthermore, presentations of an excitatory CS during extinction disrupts baseline responding when a lick-suppression measure of anticipatory behavior is used, obviating our ability to collect on-line measures of extinction after the first CS presentation of each extinction session.

**Reacclimation.** On Days 11 and 12, the subjects were reacclimated to the context as in Experiment 1.

**Testing.** On Day 13, the subjects were tested for conditioned lick suppression to X as in Experiment 1. One subject from Group OX.B died prior to the completion of the experiment. One subject from Group Con.B took longer than 60 sec to complete its initial 5 cumulative seconds of drinking prior to CS onset on the test day; therefore, the data from this subject were excluded from all statistical analyses.

**Table 2**
**Design Summary: Experiment 2**

| Group | Phase 1 | Phase 2 | Phase 3 | Test |
|---|---|---|---|---|
| OX.A | A→US/X→US/B→US | AX→US | A− | X |
| OX.B | A→US/X→US/B→US | AX→US | B− | X |
| Con.B | A→US/X→US/B→US | CX→US | B− | X |

Note—OX and Con refer to overexpectation and control treatments, respectively. Group name endings A and B refer to which stimulus was extinguished during Phase 3. A, B, and C denote the noise, tone, and buzzer stimuli, counterbalanced within groups. X denotes the click train. US represents the footshock unconditioned stimulus, and the slash (/) separates events that were interspersed.

### Results and Discussion

The central observations from Experiment 2 were the replication of the overexpectation effect (i.e., Group OX.B suppressed less to X than did Group Con.B) and recovery from the overexpectation deficit through extinction of A (i.e., stronger suppression to X in Group OX.A than in Group OX.B; see Figure 3).

A one-way ANOVA conducted on pre-CS time to complete 5 cumulative seconds of licking revealed no group differences [$F(2,67) = 1.19$, $MS_e = 0.011$, $p > .30$]. A one-way ANCOVA (with pre-CS scores as a covariate) conducted on time to complete 5 cumulative seconds of licking in the presence of X revealed an effect of treatment [$F(2,66) = 10.06$, $MS_e = 0.226$, $p < .001$]. Planned comparisons using the error term from the ANCOVA revealed lower suppression to X in Group OX.B than in Group Con.B, demonstrating overexpectation [$F(1,66) = 10.41$, $p < .005$]. Suppression to X in Group OX.A was also greater than in Group OX.B, demonstrating recovery from overexpectation through extinction of A [$F(1,66) = 18.57$, $p < .001$].

Thus, Experiment 2 showed that, as with other cue-competition effects, overexpectation can be reversed by massively extinguishing A, the putative comparator stimulus for X. However, such a demonstration fails to differentiate the performance-focused comparator hypothesis from recently formulated acquisition-focused models of retrospective revaluation effects (e.g., Dickinson & Burke, 1996; Van Hamme & Wasserman, 1994).

## EXPERIMENT 3

The purpose of Experiments 3 and 4 was to test additional predictions made by the comparator hypothesis and acquisition-focused models of retrospective revaluation effects. Blaisdell, Denniston, and Miller (1998) observed that the overshadowing deficit is strongest when both CSs share the same temporal relationship to the US and grows weaker with increasing differences in their temporal relationships to the US. Elaborating on this finding, Blaisdell, Denniston, and Miller (1999; also see Burger et al., 2000) attenuated the overshadowing deficit by shifting the overshadowing CS–US temporal interval (either from a trace to a delay relationship or vice versa) *after* overshadowing treatment, so that it differed from the overshadowed CS–US interval. The specific test we used in the present research was the same as that employed by Blaisdell, Den-

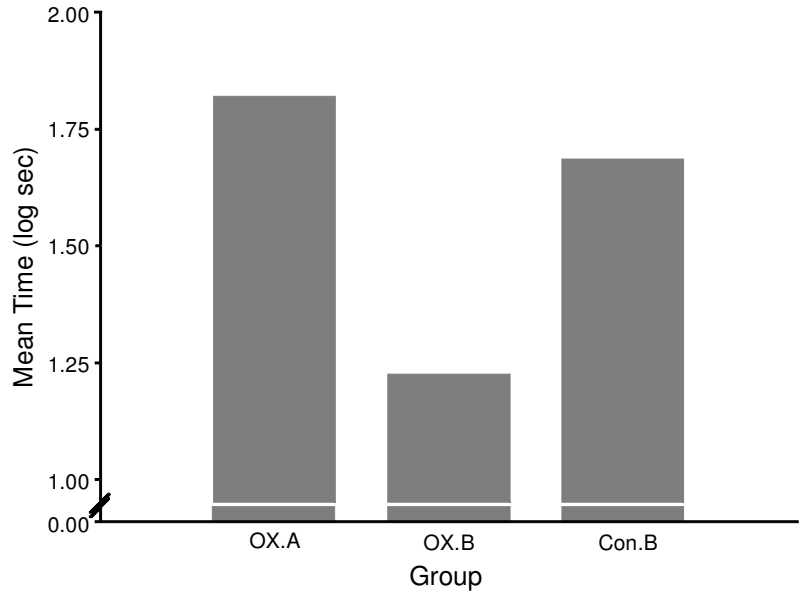


**Figure 3. Experiment 2: Adjusted mean times (log sec) to complete 5 cumulative seconds of licking in the presence of the target CS (X). For the procedure, see Table 2.**

niston, and Miller (1999) to differentiate these two families of models in how they account for overshadowing. As was previously mentioned, a comparator stimulus is effective to the extent that it provides information similar to the CS, including when the US is "expected." Thus, cue competition (e.g., overshadowing and overexpectation) should be maximal when the comparator stimulus conveys exactly the same temporal information as the target CS about the US at test. This leads to the prediction that posttraining changes in the temporal location of the US predicted by A relative to the temporal location of the US predicted by X should attenuate the effectiveness of A in attenuating responding to X. (See Blaisdell, Denniston, & Miller, 1997, for a similar statement about qualitative attributes of the US.)

Retrospective revaluation models (e.g., Dickinson & Burke, 1996; Van Hamme & Wasserman, 1994), however, are not able to fully account for the attenuation of cue-competition effects achieved by posttraining changes in the A–US temporal interval away from the value of the X–US interval. Blaisdell, Denniston, and Miller (1999) conducted posttraining shifts of the A–US interval either from a trace to a delay relationship or from a delay to a trace relationship. Retrospective revaluation models might account for the delay-to-trace effect in terms of degrading competing stimulus A's associative strength (trace conditioning has been shown to support weaker conditioned responding than does delay conditioning; e.g., Ellison, 1964; Kamin, 1954, 1965; Pavlov, 1927; but see Cole, Barnet, & Miller, 1995b). However, these models fail to account for the effect of the trace-to-delay shift in the A–US interval on overshadowing. According to Dickinson and Burke and to Van Hamme and Wasserman, shifting the A–US temporal relationship from a trace to a delay relationship should, if anything, increase A's associative strength, thereby diminishing X's associative strength. Thus, shifts of the A–US temporal relationship away from the X–US temporal relationship in both directions provide a means of contrasting these two acquisition-focused

**Table 3**
**Design Summary: Experiment 3**

| Group | Phase 1 | Phase 2 | Phase 3 | Test |
|---|---|---|---|---|
| OX.Diff | A$\overset{0}{\rightarrow}$US/X$\overset{0}{\rightarrow}$US/B$\overset{0}{\rightarrow}$US | AX$\overset{0}{\rightarrow}$US | B$\overset{5}{\rightarrow}$US | X |
| OX.Same | A$\overset{0}{\rightarrow}$US/X$\overset{0}{\rightarrow}$US/B$\overset{0}{\rightarrow}$US | AX$\overset{0}{\rightarrow}$US | A$\overset{0}{\rightarrow}$US | X |
| OX.None | A$\overset{0}{\rightarrow}$US/X$\overset{0}{\rightarrow}$US/B$\overset{0}{\rightarrow}$US | AX$\overset{0}{\rightarrow}$US | B$\overset{5}{\rightarrow}$US | X |
| Con.None | A$\overset{0}{\rightarrow}$US/X$\overset{0}{\rightarrow}$US/B$\overset{0}{\rightarrow}$US | CX$\overset{0}{\rightarrow}$US | A$\overset{0}{\rightarrow}$US | X |

Note—OX and Con refer to overexpectation and control treatments, respectively. *Same* and *Diff* indicate whether the A–US temporal interval during Phase 3 was the same as or different from, respectively, that of Phases 1 and 2. *None* indicates no Phase 3 treatment with Xs comparator stimulus. A, B, and C denote the noise, tone, and buzzer stimuli, counterbalanced within groups. X denotes the click train. US represents the footshock unconditioned stimulus, and the slash (/) separates events that were interspersed. Superscripts denote duration of gap (in seconds) between CS termination and US onset.

models with the comparator hypothesis. (In the General Discussion section, we describe how the retrospective revaluation models make these predictions.)

Consequently, in Experiments 3 and 4, we shifted (after Phase 2 of overexpectation treatment) the A–US temporal interval either from a delay to a trace relationship (Experiment 3) or from a trace to a delay relationship (Experiment 4). In Experiment 3 (see Table 3), the comparator hypothesis and the retrospective revaluation models both predicted greater responding to X when, in Phase 3, A was made a trace excitor (Group OX.Diff) than when A was maintained as a delay excitor (Group OX.Same). Responding to X by these two groups was expected to be weaker than responding by Group OX.None, which received no further conditioning of A in Phase 3.

## Method

### Subjects

Twenty-four male (300–460 g) and 24 female (200–315 g) naive, Sprague-Dawley–descended rats, bred in our colony from Holtzman stock, served as subjects. The subjects were randomly assigned to one of four groups (*n*s = 12 for Groups OX.Diff, OX.Same, OX.None, and Con.None), counterbalanced for sex. The animals were housed and maintained as in Experiments 1 and 2. The apparatus and stimuli were identical to those used in Experiment 2. The noise, tone, and buzzer served as stimuli A, B, and C, counterbalanced within groups. The clicks served as CS X for all the subjects.

### Procedure

Table 3 summarizes the critical aspects of the procedure, along with an explanation of the group names.

**Acclimation.** On Day 1, the subjects were acclimated to the context as in Experiments 1 and 2.

**Phase 1 (element training).** On Days 2 and 3, the subjects received the same Phase 1 treatment as in Experiment 2.

**Phase 2 (compound training).** On Days 4–6, the subjects in Groups OX.Diff, OX.Same, and OX.None received 6 daily AX→US trials (in the same manner as in Experiment 2), whereas the subjects in Group Con.None received 6 daily CX→US trials with a mean intertrial interval of 11 ± 4 min, for a total of 18 trials per group across the daily 60-min sessions.

**Phase 3 (shifting of A training).** On Days 7–12, the subjects in Groups OX.Diff and OX.None received four daily trace A→US or B→US trials, respectively, with a 5-sec gap between CS termination and US onset. The subjects in Groups OX.Same and Con.None received equivalent A→US pairings, but with no gap between CS termination and US onset (i.e., the A–US temporal relationship established in Phases 1 and 2 was maintained). Trials were pseudorandomly distributed in the session, with an interval of 15 ± 5 min, for a total of 24 trials across the daily 60-min sessions.

**Reacclimation.** On Days 13 and 14, the subjects were reacclimated to the context as in Experiments 1 and 2.

**Testing.** On Day 15, the subjects were tested for conditioned lick suppression to X as in Experiments 1 and 2. Two subjects met our pre-CS criterion of elimination; thus, their data were excluded from all statistical analyses.

## Results and Discussion

The central observations from Experiment 3 were the replication of the overexpectation deficit (i.e., Group OX.None suppressed less to X than did Group Con.None) and recovery from overexpectation through a posttraining shift in the A–US interval (i.e., stronger suppression to X in Group OX.Diff than in Group OX.Same; see Figure 4).

A one-way ANOVA conducted on pre-CS time to complete 5 cumulative seconds of licking revealed no group differences [$F(3,42) < 1$, $MS_e = 0.073$]. A one-way ANCOVA (with pre-CS scores as a covariate) conducted on time to complete 5 cumulative seconds of licking in the presence of X revealed an effect of treatment [$F(3,41) = 4.95$, $MS_e =$

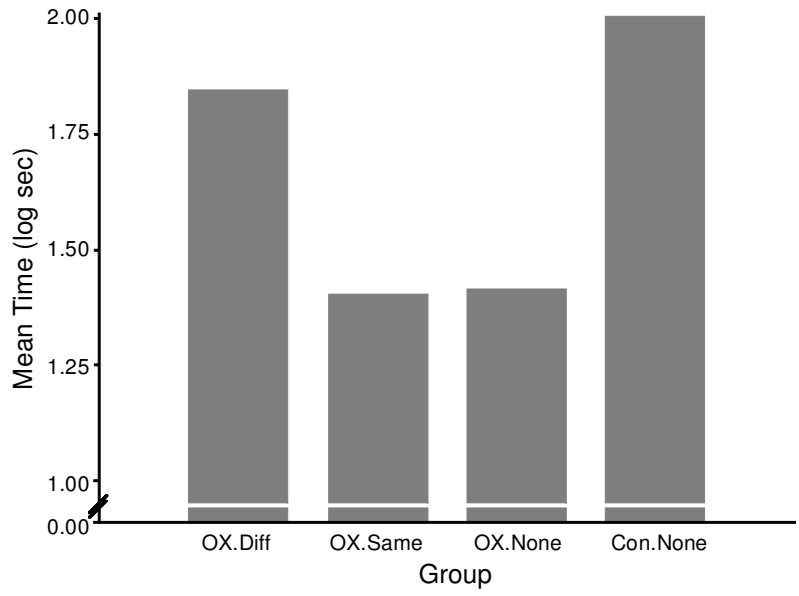


**Figure 4. Experiment 3: Adjusted mean times (log sec) to complete 5 cumulative seconds of licking in the presence of the target CS (X). For the procedure, see Table 3.**

0.211, $p < .01$]. Planned comparisons using the error term from the ANCOVA revealed less suppression to X in Group OX.None than in Group Con.None, replicating the overexpectation effect [$F(1,41) = 9.41$, $p < .005$]. Likewise, Group OX.Same suppressed less to X than did Group Con.None [$F(1,41) = 9.76$, $p < .005$], demonstrating an overexpectation effect even after delay conditioning of A in Phase 3. However, posttraining shifting of the A–US interval (Group OX.Diff) resulted in a recovery from overexpectation [Group OX.Same; $F(1,41) = 5.21$, $p < .03$]. This recovery effect was specific to posttraining shifting of A [Group OX.Diff vs. Group OX.None; $F(1,41) = 4.91$, $p < .05$] and was not observed with delay conditioning of the CS during Phase 3. Thus, consistent with the comparator hypothesis, a posttraining shift in the A–US interval from a delay to a trace relationship attenuated the overexpectation effect.

An alternative account of the increased suppression to X observed in Group OX.Diff is that associative strengths of X and the context summated during testing. That is, during the trace A→US pairings of Phase 3, the context was more contiguous with the US than was A, which might have increased conditioning of the context. Testing on X in this more excitatory context might have increased the magnitude of suppression observed at test. However, this account would have predicted a similar increase in suppression to X for Group OX.None, which received trace B→US pairings in Phase 3. Because Group OX.None evidenced little suppression to X, this alternative account becomes less plausible.

## EXPERIMENT 4

Experiment 4 used the same design as that in Experiment 3, except for the following change: During Phases 1 and 2 of overexpectation treatment, all CSs were trace conditioned, with a 5-sec gap between CS termination and US onset (see Table 4). In Experiment 4, the comparator hypothesis predicts greater suppression to CS X when (in Phase 3) A was made a delay excitor (Group OX.Diff) than when A was maintained as a trace excitor (Group OX.Same). This prediction is based on the fact that delay conditioning should have resulted in different X–US and A–US temporal relationships at the time of testing (i.e., A and X predicted the US at different times after CS onset), whereas in the trace condition CSs A and X predicted the US at the same interval after CS onset. In contrast, as will be explained in the General Discussion section, acquisition-focused models of retrospective revaluation (e.g., Dickinson & Burke, 1996; Van Hamme & Wasserman, 1994) predict no effect of a posttraining shift in the A–US temporal interval from a trace to a delay relationship.

### Method

#### Subjects

Twenty-four male (300–460 g) and 24 female (205–380 g) naive, Sprague-Dawley–descended rats, bred in our colony from Holtzman stock, served as subjects. The subjects were randomly assigned to one of four groups ($ns = 12$), counterbalanced for sex. The animals were housed and maintained as in the prior experiments. The apparatus and stimuli were identical to those used in Experiments 2 and 3. The noise, tone, and buzzer served as stimuli A, B, and C, counterbalanced within groups. The clicks served as CS X for all the subjects.

#### Procedure

Table 4 summarizes the critical aspects of the procedure, along with an explanation of the group names.

**Acclimation.** On Day 1, the subjects were acclimated to the context as in Experiments 1, 2, and 3.

**Phase 1 (element training).** On Days 2 and 3, the subjects received the same Phase 1 treatment as in Experiments 2 and 3, except for the imposition of a 5-sec gap between CS termination and US onset for all the groups.

**Phase 2 (compound training).** On Days 4–6, the subjects received the same Phase 2 treatment as in Experiment 3, except for the imposition of a 5-sec gap between CS termination and US onset for all the groups.

**Phase 3 (shifting of A training).** On Days 7–12, the subjects in Groups OX.Diff and OX.None received four daily delay A→US or B→US trials, respectively, with no gap between CS termination and US onset. The subjects in Groups OX.Same and Con.None received equivalent A→US pairings, but with a 5-sec gap between CS termination and US onset. Trials were pseudorandomly distributed in the session, with an interval of 15 ± 5 min, for a total of 24 trials across all of the daily 60-min sessions. (It is worthy of note that A→US pairings are procedurally a posttraining comparator *inflation* manipulation for Group OX.Same in this study [as well as in Experiment 3], which is predicted by the comparator hypothesis to attenuate responding to the CS. However, Denniston, Miller, and Matute, 1996, and Miller, Hallam, and Grahame, 1990, have shown that inflation effects do not occur when biologically significant stimuli are

**Table 4**
**Design Summary: Experiment 4**

| Group | Phase 1 | Phase 2 | Phase 3 | Test |
|---|---|---|---|---|
| OX.Diff | $A \xrightarrow{5} US/X \xrightarrow{5} US/B \xrightarrow{5} US$ | $AX \xrightarrow{5} US$ | $A \xrightarrow{0} US$ | X |
| OX.Same | $A \xrightarrow{5} US/X \xrightarrow{5} US/B \xrightarrow{5} US$ | $AX \xrightarrow{5} US$ | $A \xrightarrow{5} US$ | X |
| OX.None | $A \xrightarrow{5} US/X \xrightarrow{5} US/B \xrightarrow{5} US$ | $AX \xrightarrow{5} US$ | $B \xrightarrow{0} US$ | X |
| Con.None | $A \xrightarrow{5} US/X \xrightarrow{5} US/B \xrightarrow{5} US$ | $CX \xrightarrow{5} US$ | $A \xrightarrow{5} US$ | X |

Note—OX and Con refer to overexpectation and control treatments, respectively. *Same* and *Diff* indicate whether the A–US temporal interval during Phase 3 was the same as or different from, respectively, that of Phases 1 and 2. *None* indicates no Phase 3 treatment with X's comparator stimulus. A, B, and C denote the noise, tone, and buzzer stimuli, counterbalanced within groups. X denotes the click train. US represents the footshock unconditioned stimulus, and the slash (/) separates events that were interspersed. Superscripts denote duration of gap (in seconds) between CS termination and US onset.

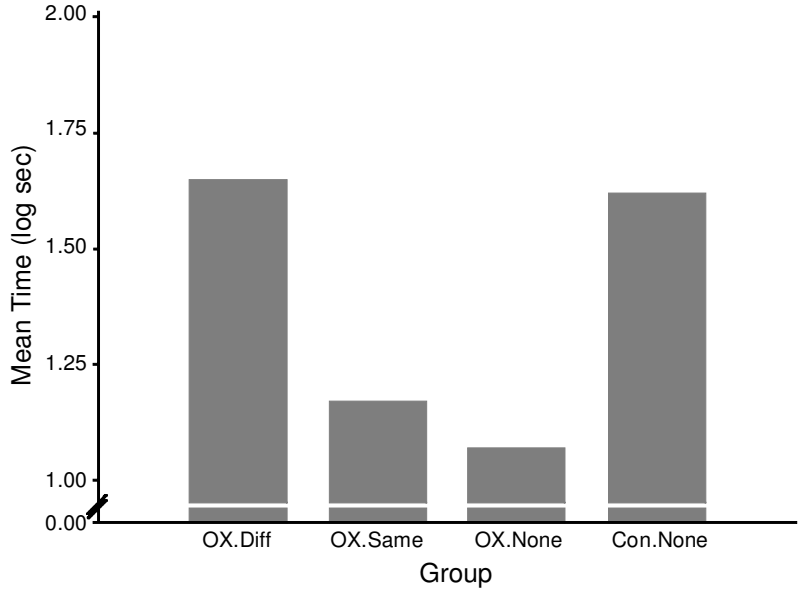


**Figure 5. Experiment 4: Adjusted mean times (log sec) to complete 5 cumulative seconds of licking in the presence of the target CS (X). For the procedure, see Table 4.**

used during initial training, as they were in this experiment [i.e., outcome = footshock]. In fact, in Experiment 3, Group OX.Same did not display decreased responding to CS X relative to Group OX.None, thereby failing to demonstrate an inflation effect when additional A→US pairings were administered. Thus, an inflation effect was not expected in this experiment. In any event, such an inflation effect in Group OX.Diff would work against our prediction of increased responding to CS X based on the requirement of temporal similarity for the directly and indirectly activated US representations.)

**Reacclimation.** On Days 13 and 14, the subjects were reacclimated to the context as in the prior experiments.

**Testing.** On Day 15, the subjects were tested for conditioned lick suppression to X as in the prior experiments. One subject (Group OX.None) failed to receive proper test treatment, owing to an equipment failure; thus, its data were excluded from all statistical analyses.

### Results and Discussion

The central observations from Experiment 4 were the replication of the overexpectation deficit (i.e., Group OX.None suppressed less to X than did Group Con.None) and recovery from overexpectation through a posttraining shift in the A–US interval (i.e., stronger suppression to X in Group OX.Diff than in Group OX.Same; see Figure 5).

A one-way ANOVA conducted on pre-CS time to complete 5 cumulative seconds of licking revealed no group differences [$F(3,43) < 1$, $MS_e = 0.057$]. A one-way ANCOVA (with pre-CS scores as a covariate) conducted on time to complete 5 cumulative seconds of licking in the presence of X revealed an effect of treatment [$F(3,42) = 7.11$, $MS_e = 0.145$, $p < .001$]. Planned comparisons using the error term from the ANCOVA revealed less suppression to X in Group OX.None than in Group Con.None, replicating the overexpectation effect [$F(1,42) = 11.70$, $p < .005$]. Likewise, Group OX.Same suppressed less to X than did Group Con.None [$F(1,42) = 8.18$, $p < .01$], demonstrating an overexpectation effect even after trace conditioning of A in Phase 3. However, posttraining shifting of A (Group OX.Diff) resulted in a recovery from overexpectation [Group OX.Same; $F(1,42) = 13.13$, $p < .001$]. Recovery was specific to posttraining shifting of A [Group OX.Diff vs. Group OX.None; $F(1,42) = 9.46$, $p < .005$], since it was not observed to result from delay conditioning of CS B during Phase 3. Thus, a posttraining shift in the A–US interval from a delay to a trace relationship attenuated the overexpectation effect.

## GENERAL DISCUSSION

Experiment 1 demonstrated overexpectation in rats, using a conditioned lick-suppression procedure, while ruling out alternative accounts in terms of US habituation and stimulus generalization decrement. Overexpectation was successfully attenuated through massive posttraining extinction of the comparator (i.e., competing) stimulus A (Experiment 2) and through posttraining shifting of the A–US temporal relationship from a delay to a trace (Experiment 3) or from a trace to a delay (Experiment 4) relationship. Thus, posttraining modification of either the strength or the temporal attributes of the A–US association reduces its effectiveness in attenuating responding to CS X.

Many contemporary models of associative acquisition can account for the basic overexpectation effect (e.g., Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972; Wagner, 1981), yet fail to adequately account for the results of Experiments 2–4. These models fail because they attribute the lack of responding produced by

overexpectation treatment to a loss of associative strength during Phase 2 compound training (i.e., unlearning of the X–US association). Once associative strength to a CS is lost, these models fail to predict recovery of responding to the target CS by any posttraining manipulation short of additional target-CS→US pairings.

Some models of associative acquisition, however, have recently been modified to explain retrospective revaluation effects, such as backward blocking (e.g., Miller & Matute, 1996; Shanks, 1985; Wasserman & Berglan, 1998), recovery from overshadowing (e.g., Kaufman & Bolles, 1981; Matzel et al., 1985), blocking (e.g., Blaisdell, Gunther, & Miller, 1999), and the relative stimulus validity effect (e.g., Cole et al., 1995a) through extinction of the competing stimulus. For example, Van Hamme and Wasserman (1994) have revised the Rescorla–Wagner (1972) model so that absent stimuli have a negative associability (in the original Rescorla–Wagner model, absent stimuli have an associability of zero). With similar consequences, Dickinson and Burke (1996) have modified Wagner's (1981) SOP so that excitatory associations can develop when two stimulus representations are simultaneously activated in the A2 state of activity and inhibitory associations can develop when one stimulus representation is activated in the A2 state and the other stimulus representation is activated in the A1 state of activity. Both of these revised models can account for the outcome of Experiment 2 (extinction of the competing cue) and possibly Experiment 3 (temporally shifting the competing cue from a delay to a trace CS). In Phase 3 of Experiment 2, A was presented alone (i.e., extinction treatment) while X and the US were both absent. Because these models assume that absent stimuli have a negative associability or enter into inhibitory associations, the X–US association should have been strengthened, thereby accounting for the increased responding to X after extinction of A. Thus, according to the Van Hamme and Wasserman model, in Phase 3 the product of the negative associability of X with the negative value of the parenthetical term (i.e., $0-V_A$) is positive, leading to increases in the strength of the X–US association. Similarly for Dickinson and Burke, in Phase 3, A should have activated both X and the US into A2, thereby promoting an excitatory X–US association.

Likewise, in Phase 3 of Experiment 3 for Group OX.Diff, further A–US trace conditioning trials occurred. Again, because absent stimuli presumably have a negative associability (in the Van Hamme & Wasserman, 1994, model) and because the US, although present, presumably had relatively poor temporal contiguity to the representation of the absent CS X, excitatory behavior to X might increase relative to Group OX.Same, for which the A–US temporal relationship was maintained between phases. However, contemporary acquisition-focused models designed to account for retrospective revaluation fail to predict the outcome of Experiment 4. In Phase 3, A was shifted closer to the US for Group OX.Diff, resulting in the representation of A's having a closer temporal contiguity to the US. Because X was physically absent, acquisition-focused models predict, if anything, a loss of conditioned responding to X (relative to Group OX.Same) as a consequence of A–US delay conditioning in Phase 3. But an increase in responding to X was observed, in direct contrast to the prediction of these models.

Of the models discussed in the introduction, only the comparator hypothesis (Denniston et al., 2001; Miller & Matzel, 1988), a performance-focused model of Pavlovian conditioning, can account for the entire data set from Experiments 1–4. In the framework of the comparator hypothesis, overexpectation results from a difference between the experimental and the control groups in the magnitude of the US representation activated indirectly by X's comparator stimulus at test (Figure 1). For example, in Group OX of Experiment 1, enough Phase 2 AX→US pairings should have resulted in A's becoming the primary comparator stimulus for X by virtue of its high temporal contiguity with X and its high saliency relative to the context (see Blaisdell, Savastano, & Miller, 1999, and Denniston et al., 2001, for elaboration of the selection conditions for comparator stimuli). For the same reason, CS B should have been X's comparator stimulus for Group Con1, whereas the context should have been X's comparator stimulus for Groups Con2 and Con3 (since there was no other stimulus present during either phase of training with X). Because A was pretrained with the US, the US representation activated indirectly by A should have been stronger than the US representation activated indirectly either by the context (Groups Con2 and Con3) or by stimulus B (which was associatively neutral going into Phase 2; Group Con1). The stronger the indirectly activated US representation was at test, the weaker should have been conditioned responding observed to the target CS. Thus, responding to X was correctly predicted to be stronger in the three control groups than in Group OX of Experiment 1.

The comparator hypothesis predicts recovery from overexpectation through extinction of X's comparator stimulus (CS A), which was observed in Experiment 2. Extinction of A presumably reduced the magnitude of the US representation indirectly activated (through the mediation by A) at test, thereby attenuating A's effectiveness in attenuating conditioned responding to X. As was discussed in the introduction, the effectiveness of a comparator stimulus is also directly related to the qualitative similarity between the US representation activated directly and indirectly by the target CS, including their temporal attributes (Blaisdell, Denniston, & Miller, 1997, 1998, 1999; Burger et al., 2000). Thus, the results of Experiments 3 and 4 can also be explained in terms of reducing the effectiveness of the A–US association in attenuating responding to X. The observed recovery from the overexpectation effect through posttraining shifts in the A–US temporal interval (making dissimilar the A–US and the X–US intervals) is analogous to the recovery from overshadowing reported by Blaisdell, Denniston, and Miller and Burger et al. to result from posttraining shifts in the temporal relationship between the overshadowing CS and the US. Thus, Experiments 3 and 4 strengthen the view that any change in the comparator-

stimulus–US association, whether it be a quantitative or a qualitative change, will alter the effectiveness of the indirectly activated US representation in modulating conditioned responding to a target CS. That is, the effectiveness of the comparator process depends not only directly on the strength of the indirectly activated representation of the US relative to the directly activated representation (quantitative difference), but also directly on the similarity of the directly activated and indirectly activated US representations in all dimensions other than degree of activation (qualitative differences). In this framework, the direct CS–US interval is encoded as part of the CS–US association constituting Link 1 of Figure 1, and the indirect CS–US interval (i.e., CS–comparator-stimulus interval summated with the comparator-stimulus–US interval) is encoded as parts of Links 2 and 3. Changes that increase the similarity between the directly activated and the indirectly activated US representations should decrease responding to the target CS; changes that decrease the similarity between the directly activated and the indirectly activated US representations should increase responding to the target CS. Thus, the posttraining manipulations of Phase 3 in Experiments 3 and 4 presumably reduced the similarity between the directly and the indirectly activated representations of the US—specifically, the times at which the US was expected on the basis of the target CS and on the basis of the comparator stimulus, thereby augmenting responding to X.[1]

Note that the shift from a delay to a trace A–US relationship in Experiment 3 may alternatively be viewed as an effect of posttraining deflation of CS A (X's comparator stimulus), because the longer A–US interval presumably not only changed the encoded interstimulus interval, but may also have weakened the A–US association (Link 3). However, the posttraining temporal shift in the opposite direction in Experiment 4 had the same effect on responding to CS X, which would not be expected if the change in responding to CS X in Experiment 3 was caused by the temporal shift's simply weakening the A–US association. A different account of the results of Experiment 4 could be provided by the acquisition-focused models of retrospective revaluation that attribute a *positive* associability to absent stimuli (e.g., Hall, 1996; Holland, 1981, 1990); however, this position fails to account for the results of Experiments 2 and 3.

Recently, Rescorla (1999) has provided independent empirical support for a performance-focused interpretation of overexpectation. In his Experiments 3A and 3B, rats received overexpectation treatment in which two auditory and two visual stimuli were paired with one of two qualitatively different, but equally valued, appetitive outcomes (sucrose and pellets; i.e., $V^1 \rightarrow O^1$, $V^2 \rightarrow O^2$, $A^1 \rightarrow O^1$, $A^2 \rightarrow O^2$, with V, A, and O being a visual CS, an auditory CS, and an outcome, respectively). Subsequently, one auditory CS received further pairings with the same outcome as that in Phase 1 (i.e., $A^2 \rightarrow O^2$), whereas the other auditory CS was paired with the same outcome as that in Phase 1 in compound with one of the pretrained visual CSs (i.e., overexpectation treatment; $V^1A^1 \rightarrow O^1$). At test, each auditory CS was separately presented while the subject was engaging in one of two instrumental responses (a chain pull or a leverpress) that was previously trained with the same outcome as the auditory CS or a different outcome. It has been shown (e.g., Colwill & Rescorla, 1988) that superimposing a Pavlovian CS on instrumental responding for the same outcome as that signaled by the CS augments the instrumental response, relative to superimposing a Pavlovian CS on instrumental responding for a different outcome. This procedure was used to assess the status of the auditory-CS–outcome association after overexpectation treatment. If the auditory CS that received compound training in Phase 2 retained an association with the outcome ($O^1$) despite a loss in Pavlovian responding (magazine entry; i.e., overexpectation), it should have augmented instrumental responding for the same outcome ($O^1$), but not for the different outcome ($O^2$). Rescorla obtained this result, suggesting that overexpectation treatment did not weaken the associative strength between the auditory CS and the outcome (i.e., the $A^1$–$O^1$ association), as would be expected on the basis of the Rescorla–Wagner (1972) model. Rather, overexpectation is likely a performance effect that leaves the S–O association intact following overexpectation treatment, thus allowing Pavlovian-to-instrumental transfer.

Lattal and Nakajima (1998) have provided a well-controlled demonstration of the overexpectation effect with appetitive stimuli in both Pavlovian and instrumental paradigms. In their introduction, they correctly point out that the overexpectation effect is similar to a posttraining inflation effect. An inflation effect is the reduction in responding to a target CS (e.g., X) through a posttraining increase in the excitatory value of the target's companion stimulus (e.g., A). Although the comparator hypothesis accounts well for both inflation and deflation (i.e., extinction) effects in theory, in practice only deflation effects have been achieved with a large degree of success. Inflation effects (e.g., backward blocking) have been much more elusive with animal subjects (e.g., Miller et al., 1990). Denniston et al. (1996; see also Miller & Matute, 1996) have proposed a solution to this quandary. They suggest that biologically significant stimuli are resistant against cue-competition effects, as well as the attenuating effects of posttraining comparator inflation (e.g., backward blocking and overexpectation). In this framework, a cue's biological significance is assessed by its potential to elicit a strong, observable response, such as stimuli that typically serve as USs in Pavlovian situations (see Blaisdell, Denniston, Savastano, & Miller, 2000, for other measures). However, a CS can acquire biological significance through pairings with another biologically significant event, such as a US or another CS (i.e., second-order conditioning). Thus, a cue that comes to control a conditioned response, thereby demonstrating biological significance, should be protected from comparator inflation effects.

Lattal and Nakajima (1998) suggested that, by modifying the comparator hypothesis, taking into account the effects of biological significance, the comparator hypothe-

sis no longer predicts the overexpectation effect. That is, Phase 1 of the overexpectation treatment establishes X as a conditioned excitor (Group Con3 of Experiment 1). Lattal and Nakajima extrapolate to conclude that, because CS X acquired biological significance in Phase 1, inflating the strength of Link 3 (Figure 1) by establishing a more strongly conditioned cue (A) as X's comparator stimulus in Phase 2 should not be able to attenuate responding to X. However, we draw attention to a fundamental difference between the inflation procedures that have failed to attenuate responding to an established conditioned excitor (e.g., Miller et al., 1990) and the procedure used to produce overexpectation. Consider backward blocking as a typical posttraining inflation procedure: After AX→US pairings in Phase 1, the blocking CS (A) is paired with the US in the absence of the target CS (X) during Phase 2. Despite this posttraining inflation of X's comparator stimulus, no attenuation in responding to X is ordinarily observed, perhaps because Link 2 (X–A) is being extinguished at the same time Link 3 (A–US) is being inflated. In contrast, A and X are paired during the AX→US trials during Phase 2 of the overexpectation treatment, thereby establishing a strong X–A association (Link 2) and maintaining the A–US association (and the X–US association) established during Phase 1. Consequently, the net increase in the magnitude of the indirectly activated US representation should be larger for overexpectation treatment (increase in Link 2 and maintenance or enhancement of a strong Link 3) than for backward blocking treatment (decrease in Link 2 and increase in Link 3). If either of these treatments could be expected to surmount the protective effects of biological significance of CS X, surely the overexpectation treatment would be more likely to do so. Blaisdell et al. (2000) demonstrated that, although it is difficult to attenuate a cue's acquired biological significance, it is possible to do so with enough training. This is consistent with our examination of the overexpectation literature, which shows that relatively many AX→US trials (as compared with most attempts to obtain backward blocking) are required to produce the effect (i.e., at least 10 and, more typically, 20 trials; e.g., Kamin & Gaioni, 1974; Khallad & Moore, 1996; Kremer, 1978; Lattal & Nakajima, 1998; Rescorla, 1970). The requirement of many AX trials to obtain an overexpectation effect is also congruent with the comparator hypothesis, because sufficient AX→US trials must be given to establish an X–A association that is stronger than the X–context association, thereby causing CS A to become a more effective comparator stimulus than is the context (see Blaisdell, Savastano, & Miller, 1999).

The present series of experiments establishes the overexpectation effect (Kremer, 1978; Rescorla, 1970) as a form of cue competition, along with blocking (Kamin, 1969), overshadowing (Pavlov, 1927), and the relative stimulus validity effect (Wagner, Logan, Haberlandt, & Price, 1968; Wasserman, 1974). Furthermore, we found that, as with other forms of cue competition, overexpectation can be reversed through posttraining extinction of the target CS's comparator (i.e., competing) stimulus, as well as through posttraining shifts in the comparator-stimulus–US interval relative to the target-CS–US interval. Finally, recovery from overexpectation through posttraining shifts in the comparator-stimulus–US interval is consistent with interpretations of cue competition and recovery therefrom as performance effects (e.g., Denniston et al., 2001; Miller & Matzel, 1988), but not with interpretation of them as acquisition-based retrospective revaluation effects (e.g., Dickinson & Burke, 1996; Van Hamme & Wasserman, 1994).

## REFERENCES


Barnet, R. C., Grahame, N. J., & Miller, R. R. (1993). Temporal encoding as a determinant of blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, **19**, 327-341.

Blaisdell, A. P., Denniston, J. C., & Miller, R. R. (1997). Unblocking with qualitative change of unconditioned stimulus. *Learning & Motivation*, **28**, 268-279.

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

Blaisdell, A. P., Denniston, J. C., & Miller, R. R. (1999). Posttraining shifts in the overshadowing stimulus–unconditioned stimulus interval alleviates the overshadowing deficit. *Journal of Experimental Psychology: Animal Behavior Processes*, **25**, 18-27.

Blaisdell, A. P., Denniston, J. C., Savastano, H. I., & Miller, R. R. (2000). Counterconditioning of an overshadowed cue attenuates overshadowing. *Journal of Experimental Psychology: Animal Behavior Processes*, **26**, 74-86.

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.

Blaisdell, A. P., Savastano, H. I., & Miller, R. R. (1999). Overshadowing of explicitly unpaired conditioned inhibition is disrupted by preexposure to the overshadowed inhibitor. *Animal Learning & Behavior*, **27**, 346-357.

Burger, D. C., Mallemat, H., & Miller, R. R. (2000). Overshadowing of subsequent events and recovery thereafter. *Quarterly Journal of Experimental Psychology*, **53B**, 149-171.

Cole, R. P., Barnet, R. C., & Miller, R. R. (1995a). Relative validity effect: Learning or performance deficit? *Journal of Experimental Psychology: Animal Behavior Processes*, **12**, 293-303.

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

Coleman, S. R., & Gormezano, I. (1971). Classical conditioning of the rabbit's (*Oryctolagus cuniculus*) nictitating membrane response under symmetrical CS–US interval shifts. *Journal of Comparative & Physiological Psychology*, **77**, 447-455.

Colwill, R. M., & Rescorla, R. A. (1988). Associations between the discriminative stimulus and the reinforcer in instrumental learning. *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 155-164.

Denniston, J. C., Miller, R. R., & Matute, H. (1996). Biological significance as a determinant of cue competition. *Psychological Science*, **7**, 325-331.

Denniston, J. C., Savastano, H. I., & Miller, R. R. (2001). Learning by contiguity, responding by relative strength: The extended comparator hypothesis. 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 revaluation of causality judgements. *Quarterly Journal of Experimental Psychology*, **49B**, 60-80.

Ellison, G. D. (1964). Differential salivary conditioning to traces. *Journal of Comparative & Physiological Psychology*, **57**, 373-380.

Hall, G. (1996). Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning. *Animal Learning & Behavior*, **24**, 233-255.

HARRIS, J. A., & WESTBROOK, R. F. (1998). Retroactive revaluation of an odor–taste association. *Animal Learning & Behavior*, **26**, 326-335.
HOLLAND, P. C. (1981). Acquisition of representation-mediated conditioned food aversion. *Learning & Motivation*, **12**, 1-12.
HOLLAND, P. C. (1990). Event representation in Pavlovian conditioning: Image and action. *Cognition*, **37**, 105-131.
HOLLAND, P. C. (1999). Overshadowing and blocking as acquisition deficits: No recovery after extinction of overshadowing or blocking cues. *Quarterly Journal of Experimental Psychology*, **52B**, 307-333.
KAMIN, L. J. (1954). Traumatic avoidance learning: The effects of CS–US interval with a trace conditioning procedure. *Journal of Comparative & Physiological Psychology*, **47**, 65-72.
KAMIN, L. J. (1965). Temporal and intensity characteristics of the conditioned stimulus. In W. F. Prokasy (Ed.), *Classical conditioning* (pp. 118-147). New York: Appleton-Century-Crofts.
KAMIN, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279-296). New York: Appleton-Century-Crofts.
KAMIN, L. J., & GAIONI, S. J. (1974). Compound conditioned emotional response conditioning with differentially salient elements in rats. *Journal of Comparative & Physiological Psychology*, **87**, 591-597.
KAUFMAN, M. A., & BOLLES, R. C. (1981). A nonassociative aspect of overshadowing. *Bulletin of the Psychonomic Society*, **18**, 318-320.
KHALLAD, Y., & MOORE, J. (1996). Blocking, unblocking, and overexpectation in autoshaping with pigeons. *Journal of the Experimental Analysis of Behavior*, **65**, 575-591.
KREMER, E. F. (1978). The Rescorla–Wagner model: Losses in associative strength in compound conditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, **4**, 22-36.
LATTAL, K. M., & NAKAJIMA, S. (1998). Overexpectation in appetitive Pavlovian and instrumental conditioning. *Animal Learning & Behavior*, **26**, 351-360.
LEVITAN, L. (1975). Tests of the Rescorla–Wagner model of Pavlovian conditioning. *Bulletin of the Psychonomic Society*, **6**, 265-268.
MACKINTOSH, N. J. (1975). A theory of selective attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, **82**, 276-298.
MATZEL, L. D., SCHACHTMAN, T. R., & MILLER, R. R. (1985). Recovery of an overshadowed association achieved by extinction of the overshadowing stimulus. *Learning & Motivation*, **16**, 398-412.
MILLER, R. R., HALLAM, S. C., & GRAHAME, N. J. (1990). Inflation of comparator stimuli following CS training. *Animal Learning & Behavior*, **18**, 434-443.
MILLER, R. R., & MATUTE, H. (1996). Biological significance in forward and backward blocking: Resolution of a discrepancy between animal conditioning and human causal judgment. *Journal of Experimental Psychology: General*, **125**, 370-386.
MILLER, R. R., & MATZEL, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 51-92). San Diego: Academic Press.
MILLER, R. R., & SCHACHTMAN, T. R. (1985). Conditioning context as an associative baseline: Implications for response generation and the nature of conditioned inhibition. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 51-88). Hillsdale, NJ: Erlbaum.
PAVLOV, I. P. (1927). *Conditioned reflexes* (G.V. Anrep, Ed. & Trans.). London: Oxford University Press.
PEARCE, J. M., & HALL, G. (1980). A model for Pavlovian conditioning: Variations in the effectiveness of conditioned but not unconditioned stimuli. *Psychological Review*, **87**, 332-352.
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.
RAUHUT, A. S., MCPHEE, J. E., DIPIETRO, N. T., & AYRES, J. J. B. (2000). Conditioned inhibition training of the competing cue after compound conditioning does not reduce cue competition. *Animal Learning & Behavior*, **28**, 92-108.
RAWSON, R. A., LEITENBERG, H., MULICK, J. A., & LEFEBVRE, M. F. (1977). Recovery of extinction responding in rats following discontinuation of reinforcement of alternative behavior: A test of two explanations. *Animal Learning & Behavior*, **5**, 415-420.
RESCORLA, R. A. (1970). Reduction in the effectiveness of reinforcement after prior excitatory conditioning. *Learning & Motivation*, **1**, 372-381.
RESCORLA, R. A. (1999). Summation and overexpectation with qualitatively different outcomes. *Animal Learning & Behavior*, **27**, 50-62.
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.
SAVASTANO, H. I., & MILLER, R. R. (1998). Time as content in Pavlovian conditioning. *Behavioural Processes*, **44**, 147-162.
SHANKS, D. R. (1985). Forward and backward blocking in human contingency judgement. *Quarterly Journal of Experimental Psychology*, **37B**, 1-21.
ST. CLAIRE-SMITH, R., & MACKINTOSH, N. J. (1974). Complete suppression to a compound CS does not block further conditioning to each element. *Canadian Journal of Psychology*, **28**, 92-101.
VAN HAMME, L. J., & WASSERMAN, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning & Motivation*, **25**, 127-151.
WAGNER, A. R. (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., LOGAN, F. A., HABERLANDT, K., & PRICE, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Comparative & Physiological Psychology*, **76**, 171-180.
WASSERMAN, E. A. (1974). Stimulus–reinforcer predictiveness and selective discrimination learning in pigeons. *Journal of Experimental Psychology*, **2**, 284-297.
WASSERMAN, E. A., & BERGLAN, L. R. (1998). Backward blocking and recovery from overshadowing in human causal judgement: The role of within-compound associations. *Quarterly Journal of Experimental Psychology*, **51B**, 121-138.
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.

## NOTE

1. It is interesting to note that, although extensive amounts of extinction are typically required to reduce the effectiveness of a comparator stimulus (e.g., Blaisdell, Gunther, & Miller, 1999), far fewer Phase 2 A–US temporal shift trials were needed here to attenuate CS A's effectiveness in attenuating responding to CS X (also see Blaisdell, Denniston, & Miller, 1999; Burger et al., 2000). However, this is consistent with the general finding that acquisition occurs in fewer trials than does extinction. For example, Rawson, Leitenberg, Mulick, and Lefebvre (1977) found that changes in an established conditioned response progressed far more rapidly with counterconditioning than with extinction treatment. Thus, it is not surprising that Phase 2 A→US pairings seemingly modified the encoded temporal interval relatively swiftly (see Coleman & Gormezano, 1971), whereas extinction of CS A results in a slower decrement in the effectiveness of the A–US association.

(Manuscript received January 16, 2001;
revision accepted for publication June 27, 2001.)