

## Protection from extinction

ROBERT A. RESCORLA

*University of Pennsylvania, Philadelphia, Pennsylvania*

The effect of the presence of a conditioned inhibitor on extinction of excitatory conditioning was studied in one magazine approach and three autoshaping experiments using rats and pigeons. In each case, the presence of an inhibitor reduced responding to an exciter during extinction but allowed substantial recovery of responding to that exciter when subsequently tested separately. Control stimuli with a history of being irrelevant to reinforcement or being nonreinforced had less of a protective effect. This constitutes a clear demonstration of protection from extinction, a phenomenon of substantial theoretical and applied importance.

Extinction is one of the most basic findings in the study of Pavlovian conditioning. It is routine to observe that a conditioned stimulus (CS) that has previously been paired with an unconditioned stimulus (US) loses its power to evoke the learned response when the US is removed. Despite years of study, we remain uncertain about the mechanisms that produce this loss in response. However, one empirical generalization of broad applicability is that the amount of extinction is a function of the magnitude of the response evoked on the trials when the CS is presented alone. The greater the number of responses during extinction, the greater the associative loss that seems to result from nonreinforcement (Rescorla, 2002).

One finding that supports this generalization is that the associative decrement produced by nonreinforcing a CS depends on the associative strength of the other stimuli that are concurrently present. For instance, both Wagner (1969) and Rescorla (2000) have reported that if an excitatory stimulus, X, is nonreinforced in the presence of another excitatory stimulus, A, subsequent performance to X is especially depressed. The concurrent presence of A enhances both the responding during X and the decrement that nonreinforcement produces in the subsequently tested responding to X. Similarly, it has been reported that accompanying X by an inhibitory A on its nonreinforced trials reduces both the response to X and the decrement it experiences. The latter phenomenon, "protection from extinction," was first described by Chorzyna (1962).

Protection from extinction is a particularly interesting case for several reasons. First, it was historically important in the development of a class of models that emphasize

the interaction among the associative strengths of elements in a compound. For instance, the Rescorla–Wagner (1972) model of conditioning proposed that the total associative strength of a compound is compared with the strength that a reinforcer can produce in order to compute an error term. That error term is then used to update the associative strengths of all the elements in the compound. The phenomenon of blocking (see, e.g., Kamin, 1968), in which the presence of an excitatory A mitigates the conditioning of X when an AX compound is reinforced, is the most well-studied empirical phenomenon supporting such models. But protection from extinction has been seen as a symmetrical case in which an inhibitory A mitigates the extinction of X when an AX compound is nonreinforced. Second, protection from extinction is often cited in discussions of applying the principles of Pavlovian conditioning to clinically relevant examples of extinction (e.g., Bouton, Mineka, & Barlow, 2001; Lovibond, Davis, & O'Flaherty, 2000). It has frequently been suggested that phobias may depend on unwanted Pavlovian fear conditioning that might be removed by extinction. However, if extinction of a fear-producing X takes place in the presence of other stimuli that have inhibitory properties, it may be less effective. Indeed, Rescorla (1968) found evidence to support Soltysik's (1963) suggestion that subject-initiated avoidance responses may develop inhibitory properties that protect fear elicitors from extinction.

However, despite its importance to basic theorizing and to clinical application, protection from extinction has received surprisingly little empirical investigation. Moreover, as Lovibond et al. (2000) note, the empirical results supporting the claim that an inhibitor protects an exciter from extinction are less than convincing. The earliest studies frequently examined nonreinforcement of an AX compound after initial A+, AX– conditioning (e.g., Chorzyna, 1962; Soltysik, 1960). Those studies found that after repeated AX– trials, responding to A alone remained high. Although this result is consistent with the possibility that an inhibitory X protects the excitatory A,

---

This research was supported National Institutes of Health Grant R01 MH 67848. Thanks to Elizabeth Harper for assistance in data collection. Correspondence concerning this article should be addressed to R. A. Rescorla, Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, PA 19104 (e-mail: rescorla@cattell.psych.upenn.edu).

it is also readily interpretable without reference to inhibition, by noting that the AX compound may be treated as a different stimulus from A alone. In that case, non-reinforcement of AX would not be expected to affect responding to A. Moreover, some studies using this procedure found no evidence for protection at all (e.g., LoLordo & Rescorla, 1966). Other studies (e.g., Soltysik, Wolfe, Nicholas, Wilson, & Garcia-Sanchez, 1983) found strong evidence for protection, but provided no evidence that X had to be inhibitory in order to provide that protection. Indeed, Kamin (1968) and Bouton (1993) have both found that a neutral X present during the nonreinforcement of an excitatory A could provide some protection for A. More recently, Lovibond et al. (2000) found protection from both an inhibitory and a neutral X.

Consequently, we lack clear evidence supporting the claim that an inhibitory stimulus is especially effective at providing protection from extinction. The present experiments were performed to search for such evidence in two contemporary Pavlovian conditioning preparations, autoshaping in pigeons and magazine approach in rats.

## EXPERIMENT 1

This experiment is an initial investigation of protection from extinction in an autoshaping preparation with pigeon subjects. The birds were given repeated presentations of three localized visual stimuli, A, B, and O, each followed by a grain reinforcer. This treatment generates directed pecking at each stimulus. The birds then received conditioned inhibition training in which O was reinforced alone but not when accompanied by a tone (T). This O+, OT- training was intended to make the tone a conditioned inhibitor. Then A and B were repeatedly nonreinforced. On these trials, A was accompanied by the inhibitory T, and B was presented alone. Finally, responding was assessed to the separately presented A and B. If the inhibitory T acted to protect A from extinction, there should be greater responding to A in that test assessment.

The logic of this design depends on successful transfer of T's inhibition from the stimulus with which it was trained to the A stimulus during extinction. Earlier work (e.g., Rescorla, 1985) had shown that this transfer could be enhanced if A has a prior history of being inhibited by some other stimulus. Consequently, prior to extinction both A and B were not only presented separately and reinforced but also presented in compound with a flashing of the house light (L) and nonreinforced. This A+, AL- and B+, BL- training should leave A and B excitatory but increase their susceptibility to inhibition by T.

## Method

### Subjects and Apparatus

The subjects were 16 female Carneau pigeons about 1 year old, originally purchased from Palmetto Pigeon Plant. They had previously participated in another autoshaping experiment that had used different stimuli. The assignment of birds to groups in the present experiment was random with respect to their previous treatments.

They were housed in pairs and maintained at 80% of their free feeding weights.

The apparatus consisted of eight identical operant chambers, each measuring 27 × 27 × 35 cm. The metal front panel of each chamber had a 5 × 5 cm food magazine in its center, located 5 cm above the wire mesh floor. A 11.8 × 14.5 cm response key, constructed of clear Lucite acrylic, was centered 9 cm above the magazine, behind a 10 × 8 cm rectangular opening in the chamber wall. Mounted behind the top edge of this response key was a set of relay contacts, the closure of which detected pecking. Located 2 mm behind the response key was a Magnavox (Model CK3923) color television. This was connected to a computer programmed to generate and display visual stimuli. A black opaque strip blocked the top half of the screen from view in four of the boxes and the bottom half of the screen in the other four boxes. Three visual stimuli could be presented in the middle of the visible portion of this screen: a yellow 5 × 23 mm bar oriented 45° from the vertical, a blue bar of the same dimensions oriented 135° from the vertical, and a "target" composed of set of black and white concentric circles, each of them 4 mm thick with an outer diameter of 20 mm.

The other three walls and ceiling of the chambers were composed of clear Plexiglas. These chambers were placed in sound- and light-attenuating shells with ventilation fans providing background noise of 62 dB re 20  $\mu$ N/m<sup>2</sup>. On the rear wall of these shells was mounted a speaker that permitted the presentation of either an 80-dB white noise (N) or an 80-dB 1800-Hz tone (T). Also on that wall was a 6-W bulb that was continuously illuminated during the session, except during the operation of the food hopper. That hopper contained Purina Pigeon grain. The illumination of that bulb could also be interrupted at a rate of 1/sec to provide a diffuse visual stimulus (L).

Experimental events were automatically controlled by relay equipment and microprocessors located in an adjoining room.

### Procedure

**Conditioning Phase 1.** On each of the 4 days of this phase, the birds received Pavlovian conditioning of the form O+, A+, B+, in which "+" indicates a grain reinforcer. Each session contained twelve 5-sec presentations of each stimulus, each terminating in 5 sec of grain. For all animals, the target played the role of O and the blue and yellow bars were counterbalanced as A and B. In this and all subsequent phases of the experiment, the intertrial interval (ITI) was variable around a mean of 1 min.

**Conditioning Phase 2.** On each of the next 16 days, all animals received conditioned inhibition training of the form O+, OT-, A+, B+, AL-, BL-. Each session contained twelve 5-sec presentations of each stimulus, given in pseudorandom order with the constraint that no more than two consecutive trials were of the same type. On compound trials, the L and T were presented simultaneously with the localized visual stimuli

**Extinction.** On each of the next 2 days, the animals received nonreinforced presentations of the blue and yellow bars. Each session contained 12 intermixed presentations of O+, AT-, and B-. In addition, the session began with 2 presentations of O+ in order to minimize the disruption in performance frequently seen at the beginning of a session. For half the animals, the blue bar played the role of A and the yellow bar played the role of B; for the other half of the animals, the blue and yellow bars were interchanged.

**Test.** On the next day, the animals received nonreinforced test presentations of A and B alone. The session began with a half-session of extinction, followed without interruption by two nonreinforced presentations each of A and B, presented in a counterbalanced abba order. A single reinforced presentation of T intervened between the second and third test trials.

## Results and Discussion

Pavlovian conditioning proceeded without incident. By the final day of conditioning, the mean responses per

minute were 168.9 and 12.8 during the O+ and OT- trials, respectively. Responding to the A and B stimuli was 159.1 and 6.8 during their separate presentation and during their presentation in compound with L, respectively.

Figure 1 shows responding during the course of extinction and in the final test. It is clear that there was excellent transfer of the inhibitory power of the tone to the A stimulus. Throughout extinction, there was markedly less responding during the AT compound than during B alone. For extinction as a whole, the difference was highly reliable [Wilcoxon  $T(16) = 0, p < .01$ ]. Of more interest, there was substantially more responding to A than to B during the test session [ $T(16) = 9, p < .01$ ]. That is, the presence of T during the nonreinforcement of A produced substantial protection of A from extinction, relative to a B stimulus' receiving the same number of nonreinforced presentations in the absence of T.

Although this experiment provides evidence for protection of an autoshaped response from extinction because of the presence of T, it yields little information about the properties of T that make it effective. On the one hand, the results are consistent with the view that T's conditioned inhibition is the essential factor. That inhibition has the effect of both reducing responding during A and minimizing the error resulting from its nonreinforcement. On the other hand, it is quite possible that T would have had either or both of these effects even in the absence of its being trained as an inhibitor. Experiments 2 and 3 addressed the importance of such training.

## EXPERIMENT 2

The goal of Experiments 2 and 3 was to replicate the results of Experiment 1 under conditions that allowed evaluation of the role of conditioned inhibition in producing protection. For this purpose, two localized visual

stimuli were extinguished, one in the presence of an inhibitor and one in the presence of a control stimulus, which was less inhibitory. Experiments 2 and 3 generated that control stimulus in two different ways.

In Experiment 2, the birds were given an O+, OX-, OY+ procedure, intended to establish X as a conditioned inhibitor (signaling the omission of O's reinforcer) while arranging for Y to be neutral (not affecting the reinforcement of O). This treatment of Y is often used as a control procedure in studies of modulation (e.g., Holland, 1995; Rescorla, 1985). Its intention is to generate a control stimulus which, although presented in compound with O, is not informative with regard to either reinforcement or nonreinforcement.

## Method

### Subjects and Apparatus

The subjects were 16 birds of the same type and with histories similar to those of the birds used in Experiment 1. The apparatus was that of Experiment 1.

### Procedure

**Initial conditioning.** On the first 4 days, the birds received a treatment identical to that of initial conditioning in Experiment 1. All birds received 12 reinforced presentations each of the target, the blue bar, and the yellow bar. The order of trials was pseudorandom and the mean ITI in this and all subsequent phases was 1 min.

**Conditioned inhibition.** On each of the next 32 days, the birds received a treatment intended to establish either N or T as a conditioned inhibitor while arranging for the other to be uninformative with regard to reinforcement. On each day, they received 12 presentations each of O+, OX-, and OY+. For half the animals, the N played the role of X and T that of Y; for the other half of the animals, the roles of N and T were interchanged. For all animals, the target played the role of O+. In addition, on each day the birds received 12 intermixed reinforced presentations of the each of the two colored bars (A+ and B+) and 12 nonreinforced presentations of each colored bar presented in compound with the flashing house light (AL- and BL-). As in Experiment 1, the intention was to

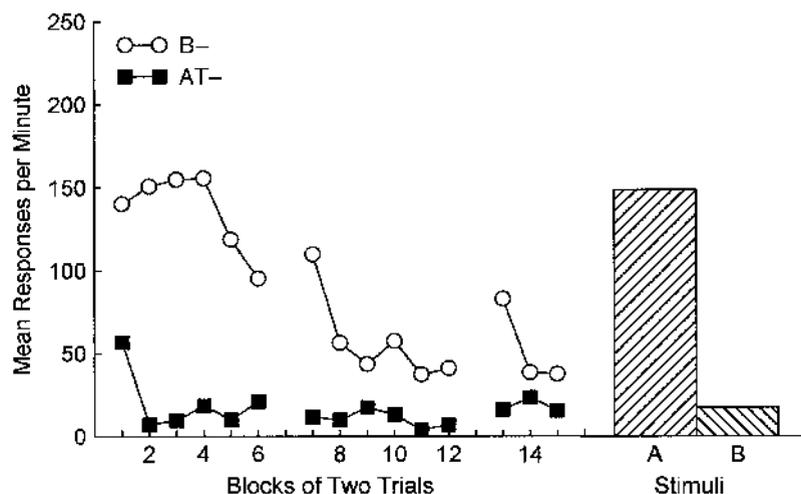


Figure 1. Mean responses per minute during extinction and testing in Experiment 1. Extinction responding is shown during the nonreinforced presentation of two visual stimuli (A and B), one of which was accompanied by an inhibitory tone (T). Test responding is shown to the separately presented A and B stimuli.

make the colored bars excitatory but vulnerable to any inhibition from X.

**Extinction.** On each of the 2 days, the animals received nonreinforced presentations of the blue and yellow bars. Each session contained 12 intermixed presentations of O+, AX-, and BY-. As in Experiment 1, the session began with 2 presentations of O+ in order to minimize the disruption in performance frequently seen at the beginning of the session. For half the animals, the blue bar played the role of A and the yellow bar played the role of B; for the other half of the animals, the blue and yellow bars were interchanged.

**Test.** On the next day, the animals received nonreinforced test presentations of A and B alone. The session began with a half-session of extinction, followed without interruption by two nonreinforced presentations each of A and B, presented in a counterbalanced abba order. A single reinforced presentation of O intervened between the second and third test trials.

### Results and Discussion

The results were highly similar to those of Experiment 1. On the final day of conditioned inhibition training, the mean responses per minute were 212.5, 37.3, and 203.1 during the reinforced target alone trials, the trials on which the target was compounded with the inhibitor, and the trials on which the target was compounded with the control stimulus, respectively. In that session, the mean responses per minute were 196.0 and 13.4 during the colored bars presented alone and in compound with the light, respectively.

Figure 2 shows the results of extinction and testing. It is clear that during extinction there was substantially more responding to the colored bar presented in compound with the control stimulus (BY) than to that presented in compound with the inhibitor (AX). Over the course of extinction, the difference was highly reliable [ $T(16) = 5, p < .01$ ]. Of more interest, during the test, there was greater responding to the bar that had been

nonreinforced in the presence of the inhibitor (A) than to that nonreinforced in the presence of the control stimulus (B) [ $T(13) = 2.5, p < .01$ ].

These results confirm those of Experiment 1, that nonreinforcement in the presence of an inhibitor can protect a stimulus from extinction. Moreover, they suggest that the inhibitor is more successful at providing this protection than is another stimulus which has been arranged to be irrelevant to reinforcement.

### EXPERIMENT 3

Experiment 2 provided one kind of control stimulus against which to evaluate the importance of inhibition for providing protection. However, despite the frequent use of the O+, OY+ treatment as a control in studies of modulation, one might be concerned that it does not in fact leave the Y stimulus neutral. It seems possible that a Y would emerge from this procedure having a slight excitatory value. Moreover, there is evidence that presenting an excitatory stimulus in compound with an extinguishing stimulus can enhance that extinction (see, e.g., Rescorla, 2000; Wagner, 1969). This makes available an alternative interpretation of the results of Experiment 2 in terms of the possible excitatory power of the control stimulus promoting extinction, rather than the inhibitory power of the conditioned inhibitor diminishing extinction.

The goal of Experiment 3 was to provide an alternative comparison stimulus that was unlikely to have any excitatory power. This was accomplished by replacing the OY+ trials of Experiment 2 with Y- trials. Consequently, initial conditioning involved O+, OX-, and Y- trials. Then X and Y were presented in compound with localized visual stimuli, A and B, at a time when they were nonreinforced. This separate nonreinforced pre-

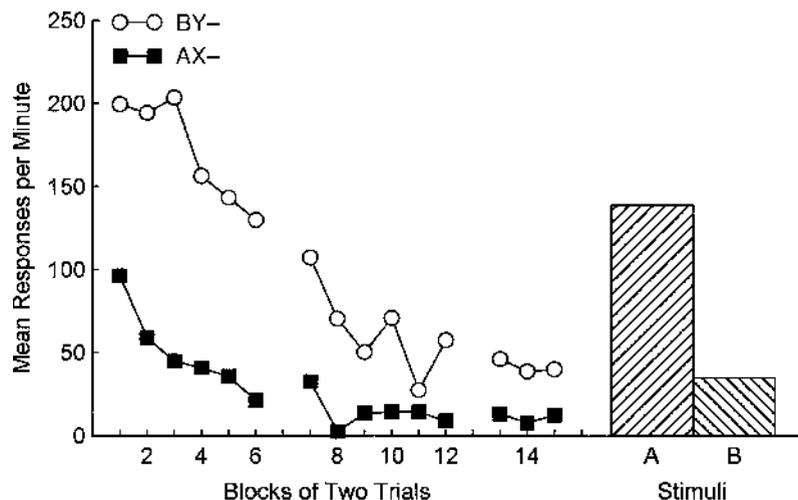


Figure 2. Mean responses per minute during extinction and testing in Experiment 2. Extinction responding is shown during the nonreinforced presentation of two visual stimuli (A and B), one accompanied by an inhibitory X and one by a control Y. Test responding is shown to the separately presented A and B stimuli.

sensation of Y should not establish it as an exciter; indeed, it has sometimes been thought that such a procedure would generate a moderate amount of inhibition to Y (e.g., Rescorla, 1969). Aside from this replacement of OY+ trials with Y- trials, the experiment was identical to Experiment 2.

### Method

#### Subjects and Apparatus

The subjects were 16 birds of the same type and with histories similar to those of the birds used in Experiment 1. The apparatus was that of Experiment 1.

#### Procedure

The procedure was identical to that of Experiment 2, except that wherever Experiment 2 arranged OY+ trials, Experiment 3 arranged Y- trials.

#### Results and Discussion

Initial conditioned inhibition training proceeded smoothly. On the final day of that training, the mean responses per minute were 197.1, 11.7, and 1.4 for the target alone (O), the target in compound with the inhibitor (OX), and with the control stimulus (Y), respectively. In that session, the mean responses per minute were 162.8 and 2.0 during the colored bars when presented alone (A and B) and in compound with the light (L), respectively.

Figure 3 shows responding over the course of extinction and test. As in Experiment 2, there was greater extinction responding during the compound of the control stimulus and the target (BY) than during the compound of the inhibitor and the target (AX) [ $T(14) = 14, p < .05$ ]. Moreover, in the test, there was greater responding to colored bars that had been extinguished in the presence of the inhibitor (A) [ $T(12) = 12, p < .05$ ].

The results of Experiment 3 were quite similar to those of Experiment 2, although the effects were numerically more modest. This somewhat weaker effect seems likely to be attributable to stimulus generalization between the inhibitor and the control stimulus. Nevertheless, in both experiments, an inhibitor provided better protection than did a control stimulus. Since it seems possible that the control stimulus used in Experiment 2 was slightly excitatory whereas that used in Experiment 3 was slightly inhibitory, the combination of the two experiments strongly suggests that the greater conditioned inhibition controlled by the diffuse stimulus is responsible for the greater protection from extinction.

### EXPERIMENT 4

This experiment was a systematic replication of Experiment 2, using a different conditioning preparation: magazine approach in rats. In this commonly used preparation, diffuse auditory and visual stimuli signal the availability of food in a particular location, the magazine. As the rats learn this relation, they increasingly approach and put their heads into the food magazine, a response that is easily detected by the interruption of a photo beam.

As in Experiment 2, the animals first received a conditioning procedure of the form O+, OX-, OY+, A+, B+, in which O, A, and B were auditory stimuli and X and Y were counterbalanced visual stimuli. This was intended to give X inhibitory properties, because it signalled nonreinforcement at the same time as it left Y relatively neutral because it signalled neither reinforcement nor nonreinforcement. Then A and B were extinguished, one in the presence of X and one in the presence of Y.

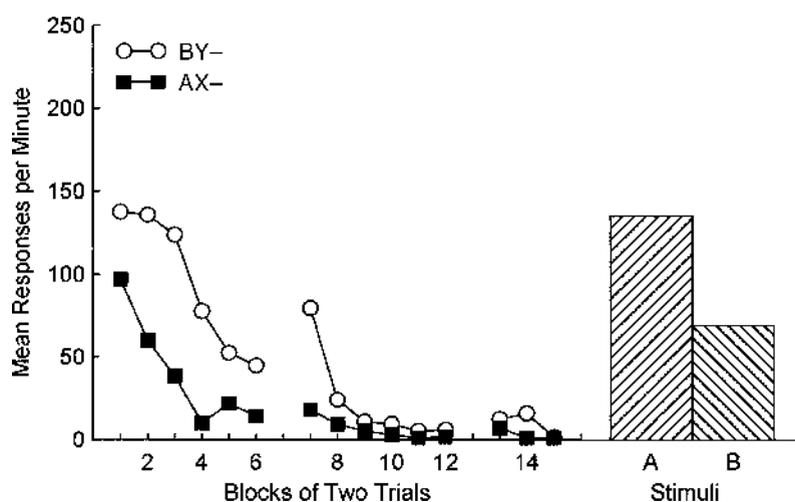


Figure 3. Mean responses per minute during extinction and testing in Experiment 3. Extinction responding is shown during the nonreinforced presentation of two visual stimuli (A and B), one accompanied by an inhibitory X and one by a control Y. Test responding is shown to the separately presented A and B stimuli.

The question of interest was whether the inhibitory X would provide superior protection from extinction of the stimulus nonreinforced in its presence.

## Method

### Subjects and Apparatus

The subjects were 16 male Sprague–Dawley rats about 90 days old. They were housed in individual cages and placed on a food deprivation regime that kept them at 80% of their ad-lib body weight for the course of the experiment. They had free access to water in the home cage.

The apparatus consisted of eight operant chambers measuring  $22.9 \times 20.3 \times 20.3$  cm, identical to those used in previous reports (e.g., Colwill & Rescorla, 1985). The two end walls of each chamber were aluminum; the sidewalls and ceiling were clear Plexiglas. The floor of the chamber was composed of 0.48-cm stainless steel rods, spaced 1.9 cm apart, center to center. Each chamber had a recessed food magazine located behind a  $4 \times 4$  cm opening in the center of one end wall. An infrared detector and emitter system was mounted on the side walls of the magazine, 1.5 cm behind the opening and 1.5 cm from the floor of the magazine, permitting automatic recording of head movements into the magazine.

Each chamber was enclosed in a sound- and light-resistant shell. Mounted on the inside wall of this shell were two speakers, one of which permitted the presentation of a white noise (N) and a 2/sec clicker (C), and the other of which permitted the presentation of a 1800-Hz tone (T), each measuring approximately 76 dB re  $20 \mu\text{N}/\text{m}^2$  against a background level of 62 dB. Also mounted on that wall was a 6-W bulb that could be illuminated to provide a light (L) stimulus during the otherwise dark session. Another 6-W light was mounted outside the rear wall of the chamber, at grid level; this light could be flashed on at a rate of 1/sec to produce a flashing (F) stimulus. A food dispenser containing 45-mg pellets (P. J. Noyes, Formula A) allowed delivery of pellets into the food magazine.

Experimental events were controlled and recorded automatically by relays and microprocessors located in an adjoining room.

### Procedure

**Magazine training.** On the 1st day, the animals received a 20-min magazine training session, during which 20 noncontingent deliveries of pellets were given, at time intervals variable around a mean of 1 min.

**Conditioning.** On each of the next 4 days, all animals received Pavlovian conditioning intended to make the three auditory stimuli excitatory. On each day, the rats received eight 30-sec presentations each of T, N, and C. Each presentation terminated in the delivery of a 45-mg pellet. On each of the next 6 days, this treatment continued but was augmented by the addition of 8 presentations each of T in compound with L and in compound with F. For half the animals, the TL compound was nonreinforced and the TF compound was reinforced; for the other half of the animals, the reinforcement contingencies were interchanged. Stimuli were presented in a balanced order, with the constraint that no more than two trials of the same type occurred consecutively. The ITI, defined from trial onset to trial onset, was variable around a mean of 2.5 min in these and all subsequent sessions. The number of photo beam interruptions by head entry into the magazine was recorded during presentation of each 30-sec stimulus and during the 30-sec stimulus-free period preceding each stimulus.

**Extinction.** On each of the next 4 days, the animals received eight nonreinforced presentations of N and of C, mixed with eight reinforced presentations of T. For half the animals, the presentations of N were accompanied by L and those of C were accompanied by F; for the remaining half of the animals, N was accompanied by F and C by L.

**Test.** On the next day, the animals received a half-session of extinction, followed without interruption by two nonreinforced test presentations each with N and C mixed with two reinforced presentations of T. The order of test trials was either NTCCTN or CTNNTC, counterbalanced with regard to the prior treatment of the animals.

### Results and Discussion

On the final day of conditioned inhibition training, the mean responses per minute were 3.8, 12.4, 2.8, and 12.0 during the prestimulus period, the T+ trials, the nonreinforced trials of T in compound with either L or F, and the reinforced trials of T in compound with F or L, respectively. The reduction in responding during the inhibitory compound, relative to that during T alone, suggests that inhibition had developed. The similarity of responding observed during T in compound with the control stimulus, relative to that during T alone, is consistent with the expectation that the control stimulus becomes relatively neutral. In that session, the mean responses per minute during the N and C were 15.8.

Figure 4 shows responding to the N and C stimuli over the course of extinction and during the test. Since there were no reliable differences as a function of stimulus identity, the data are presented separately for the auditory stimulus being extinguished in compound with the inhibitory light (AX) and that presented in compound with the control light (BY). There was reliably greater responding to the latter compound over the course of extinction [ $T(16) = 2, p < .01$ ]. Responding to the N and C alone during the test session is shown to the right. It is clear that during this test there was greater responding to the stimulus that had been nonreinforced in compound with the inhibitor (A) than that which had been nonreinforced in compound with the control stimulus (B) [ $T(15) = 25, p < .05$ ]. That is, the inhibitor provided some protection against the extinguishing effects of non-reinforcement.

## GENERAL DISCUSSION

These experiments provide clear evidence that arranging for an inhibitor to be present during the nonreinforcement of an exciter affords some protection from extinction for that exciter. This observation was made in two different conditioning preparations and under circumstances in which the effects of the inhibitor were compared with those of no stimulus, a stimulus that was irrelevant to the delivery of a reinforcer, and a stimulus that had been repeatedly presented without reinforcement.

Although these experiments demonstrate that an inhibitory stimulus is especially effective at protecting against the effects of nonreinforcement, they do not allow us to determine whether that inhibition is necessary for protection. It may well be, as Kamin (1968) found, that the simple presence of a new stimulus during extinction would provide some measure of protection. The moderate test responding to the stimulus extin-

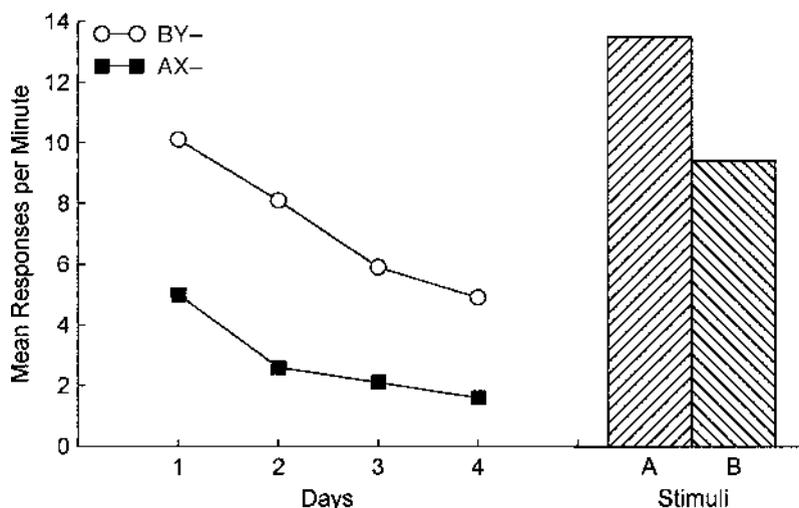


Figure 4. Mean responses per minute during extinction and testing in Experiment 4. Extinction responding is shown during the nonreinforced presentation of two visual stimuli (A and B), one accompanied by an inhibitory X and one by a control Y. Test responding is shown to the separately presented A and B stimuli.

guished in compound with the control stimulus observed in Experiments 3 and 4 is consistent with the possibility. Of course, it may well be that an initially neutral stimulus develops inhibition when it is present during the extinction of an exciter (see, e.g., Rescorla, 1999), thereby enhancing its ability to protect that exciter. Nor do these experiments allow us to decide whether or not the protection provided by an inhibitor is complete. Although test responding was substantial to the stimulus nonreinforced in the presence of the inhibitor, these experiments did not make comparison with a stimulus not subjected to extinction at all.

In these experiments, several different procedures were used for generating a control stimulus with which the inhibitor was to be compared. Which procedure seems attractive may depend on one's choice of a theory of conditioning. For instance, the Rescorla-Wagner model views both the  $O+$ ,  $OY+$  and the  $O+$ ,  $Y-$  procedures as generating a Y that is neutral. Moreover, within that theory, neither Y would be expected to initially affect responding to an excitatory A with which it is presented in compound. However, both the configural model suggested by Pearce (1987) and the hybrid model recently described by Wagner and Brandon (2001) anticipate that the  $O+$ ,  $OY+$  procedure would leave Y somewhat excitatory, whereas the  $Y-$  procedure would leave Y neutral, as assessed by responding during separate presentation of Y. Nevertheless, both those theories also anticipate that combining either of those Ys with an excitatory A during extinction should yield a decrease in responding relative to that observed during A alone. This results because of generalization decrement from A to AY. By that measure, both theories suggest that either Y would have an inhibitory impact on A. In addition, it has

frequently been suggested (e.g., Konorski, 1948; Rescorla, 1969) that simple  $O+$ ,  $Y-$  treatment could generate an inhibitor. Consequently, theoretical accounts differ on the status of Y treated in various ways as measured in various situations. What all of these accounts agree on is that the  $O+$ ,  $OY-$  procedure should in fact generate an X that is more inhibitory than either of those ways of treating Y, as assessed both when X and Y are tested alone and when they are presented in compound with A. That is, although there is no agreement on the degree to which a procedure generates a truly neutral Y, there is agreement on the relative amounts of inhibition controlled by X and Y. Under those circumstances, the wisest procedure is to compare the effects of X on protecting A with several different Y treatments. The conclusion which these comparisons permit is that the more inhibitory X has a better protecting effect.

The phenomenon of protection from extinction is relevant to a variety of empirical, theoretical, and practical issues. Empirically, protection is quite similar to the phenomenon of "renewal" after extinction. Bouton and his colleagues have reported that a stimulus that is extinguished in one context will again generate substantial responding if tested in a different context (e.g., Bouton & Brooks, 1993). It has been suggested that the extinction context serves to protect the original association, which then reveals itself in the different test context. These renewal results differ from those reported here in two ways. First, they involve the effects of a contextual, rather than a discrete, stimulus. It is not uncommon to view contextual stimuli as playing a different role than discrete stimuli in learning situations. Second, the contextual stimuli are initially neutral, rather than inhibitory. As in the case of a discrete stimulus, it seems possible

that over the course of extinction a context becomes inhibitory because of the nonreinforcement of an exciter in its presence. Renewal might then occur because this inhibitor is removed during the test in a different context. However, Bouton (1991) has argued against that interpretation, because of the stimulus specificity of the control exerted by the context. Instead, he has suggested that contexts might have their renewal effect because they form a configuration with the stimulus or they develop conditional control.

Protection from extinction is also concordant with the empirical generalization that the magnitude of responding in extinction affects the impact of nonreinforcement. Rescorla (2002) has noted that a number of manipulations enhance or diminish responding in extinction and also enhance or diminish the associative change that results. For instance, if extinction is conducted at a time when the animal is satiated, the stimulus evokes a reduced response but also undergoes less associative change. Moreover, there is evidence that if extinction is conducted in the presence of another exciter, so as to augment the responding, this augments the change produced by nonreinforcement (Rescorla, 2000; Wagner, 1969). This may be thought of as the mirror image of protection in which the presence of an inhibitor diminishes both responding and the associative change. All of these results have encouraged the view that extinction may importantly involve learning about responding (e.g., Rescorla, 2002).

Theoretically, protection from extinction has been viewed as supporting a class of error-correction models of conditioning (e.g., Rescorla & Wagner, 1972). According to such models, associative change is governed by the discrepancy between the current associative strength of a compound and that which the consequent reinforcer or nonreinforcer can produce. If an exciter is accompanied either by another exciter or by an inhibitor, this should augment or diminish the total associative strength, respectively. That in turn should modulate the ability of a reinforcer to increase or a nonreinforcer to decrease associative strength. The most well-studied example, of course, is blocking, in which the presence of another exciter diminishes the reinforcing power of a US. Protection from extinction has the same theoretical implication.

Of course, it is of interest to ask whether one could separate the effects of an inhibitor on responding from those on the net underlying associative strength. But this is likely to prove quite difficult, because reduction in responding is the clearest indication that the associative strength has been reduced. If one should present a conditioned inhibitor and find that it failed to reduce responding during extinction, one would be inclined to conclude that the inhibition did not transfer successfully and that therefore the net associative strength was not in fact reduced. On the other hand, if one should present a stimulus deemed not to be an inhibitor and it did reduce responding, one might be inclined to suppose that the stimulus

had reduced the effective associative strength, perhaps by modifying the processing of the stimulus.

The Pavlovian phenomenon of protection from extinction has been used as an explanatory concept in various other settings. For instance, it has been suggested that it may help account for the frequent observation that instrumental avoidance behavior is resistant to extinction. Soltysik (1963) has noted that the instrumental response in avoidance learning situations has a relation to the fear-evoking signal which is logically like that of a Pavlovian inhibitor: The signal terminates in shock except when it is accompanied by the response. Rescorla (1968) has provided evidence that an external stimulus that accompanies an avoidance response indeed develops conditioned inhibition. This supports the suggestion made by Soltysik that the occurrence of an avoidance response might result in the avoidance signal's being accompanied by an inhibitor on successful avoidance trials. That in turn might help explain the observation that avoidance behavior can be difficult to extinguish. The inhibitory properties of the avoidance response itself serves to protect the avoidance signal from the decremental effects of nonreinforcement. This in turn might mean that natural avoidance behavior that a human makes in the presence of a phobic object could help to preserve the phobia.

Finally, protection from extinction has been seen as having important application to certain clinical phenomena. To the degree that one believes that changes in emotional reactions to a stimulus might be governed by Pavlovian mechanisms, extinction might be an important clinical tool. However, the ability to change emotional reactions by extinction may be compromised if the change procedure makes available a conditioned inhibitor. Some authors (e.g., Bouton et al., 2001; Lovibond et al., 2000) have argued that in therapeutic settings there may be a variety of stimuli, including the therapist herself, which might serve as an inhibitor of fear responses and thus diminish the ability of extinction to change the value of a stimulus.

For all of these reasons, the phenomenon of protection from extinction has received frequent mention. Perhaps because it accords so well with a variety of other ideas, it has not previously received the kind of careful empirical documentation that it deserves. The present experiments make a contribution toward that documentation.

## REFERENCES

- BOUTON, M. E. (1991). Context and retrieval in extinction and in other examples of interference in simple associative learning. In L. W. Dachowski & C. F. Flaherty (Eds.), *Current topics in animal learning: Brain, emotion, and cognition* (pp. 25-53). Hillsdale, NJ: Erlbaum.
- BOUTON, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, **114**, 8-99.
- BOUTON, M. E., & BROOKS, D. C. (1993). Time and context effects on performance in a Pavlovian discrimination reversal. *Journal of Experimental Psychology: Animal Behavior Processes*, **19**, 165-179.
- BOUTON, M. E., MINEKA, S., & BARLOW, D. H. (2001). A modern learn-

- ing theory perspective on the etiology of panic disorder. *Psychological Review*, **108**, 4-32.
- CHORAZYNA, H. (1962). Some properties of conditioned inhibition. *Acta Biologiae Experimentalis*, **22**, 5-12.
- COLWILL, R. M., & RESCORLA, R. A. (1985). Postconditioning devaluation of a reinforcer affects instrumental responding. *Journal of Experimental Psychology: Animal Behavior Processes*, **11**, 120-132.
- HOLLAND, P. C. (1995). Transfer of occasion setting across stimulus and response in operant feature positive discrimination. *Learning & Motivation*, **26**, 239-263.
- KAMIN, L. J. (1968). Attention-like processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior: Aversive stimuli* (pp. 9-32). Coral Gables, FL: University of Miami Press.
- KONORSKI, J. (1948). *Conditioned reflexes and neuron organization*. Cambridge: Cambridge University Press.
- LOLORDO, V. M., & RESCORLA, R. A. (1966). Protection of the fear-eliciting stimulus from extinction. *Acta Biologiae Experimentalis*, **26**, 251-258.
- LOVIBOND, P. F., DAVIS, N. R., & O'FLAHERTY, A. S. (2000). Protection from extinction in human fear conditioning. *Behaviour Research & Therapy*, **38**, 967-983.
- PEARCE, J. M. (1987). A model of stimulus generalization for Pavlovian conditioning. *Psychological Review*, **84**, 61-73.
- RESCORLA, R. A. (1968). Pavlovian fear conditioning in Sidman avoidance learning. *Journal of Comparative & Physiological Psychology*, **65**, 55-60.
- RESCORLA, R. A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin*, **72**, 77-94.
- RESCORLA, R. A. (1985). Conditioned inhibition and facilitation. In R. R. Miller & N. S. Spear (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 299-326). Hillsdale, NJ: Erlbaum.
- RESCORLA, R. A. (1999). Partial reinforcement reduces the associative change produced by nonreinforcement. *Journal of Experimental Psychology: Animal Behavior Processes*, **25**, 403-414.
- RESCORLA, R. A. (2000). Extinction can be enhanced by a concurrent exciter. *Journal of Experimental Psychology: Animal Behavior Processes*, **26**, 251-260.
- RESCORLA, R. A. (2002). Extinction. In L. Bäckman & C. von Hofsten (Eds.), *Psychology at the turn of the millennium: Vol. 1. Cognitive, biological, and health perspectives* (pp. 217-244). Hove, U.K.: Taylor & Francis.
- RESCORLA, R. A., & WAGNER, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.
- SOLTYSIK, S. [S.] (1960). Studies on the avoidance conditioning: III. Alimentary conditioned reflex model of the avoidance reflex. *Acta Biologiae Experimentalis*, **20**, 183-192.
- SOLTYSIK, S. [S.] (1963). Inhibitory feedback in avoidance conditioning. *Boletín del Instituto de Estudios Médicos y Biológicos México*, **21**, 434-448.
- SOLTYSIK, S. S., WÖLFE, G. E., NICHOLAS, T., WILSON, J., & GARCIA-SANCHEZ, J. L. (1983). Blocking of inhibitory conditioning within a serial conditioned stimulus-conditioned inhibitor compound: Maintenance of acquired behavior without an unconditioned stimulus. *Learning & Motivation*, **14**, 1-29.
- WAGNER, A. R. (1969). Stimulus selection and a "modified continuity theory." In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 3, pp. 1-41). New York: Academic Press.
- WAGNER, A. R., & BRANDON, S. E. (2001). A componential theory of Pavlovian conditioning. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 23-64). Mahwah, NJ: Erlbaum.

(Manuscript received October 15, 2002;  
revision accepted for publication January 27, 2003.)