Research

AX+, BX– Discrimination Learning in the Fear-Potentiated Startle Paradigm: Possible Relevance to Inhibitory Fear Learning in Extinction

Karyn M. Myers[1,3] and Michael Davis[1,2,3,4]

[1]Departments of Psychology and [2]Psychiatry and Behavioral Sciences, [3]Center for Behavioral Neuroscience, Emory University, Atlanta, Georgia 30322, USA

The neural mechanisms of fear suppression most commonly are studied through the use of extinction, a behavioral procedure in which a feared stimulus (i.e., one previously paired with shock) is nonreinforced repeatedly, leading to a reduction or elimination of the fear response. Although extinction is perhaps the most convenient index of fear inhibition, a great deal of behavioral work suggests that postextinction training conditioned stimuli are both excitatory and inhibitory, making it difficult to determine whether a neural manipulation affects inhibition, excitation, or some combination thereof. For this reason we sought to develop a behavioral procedure that would render a stimulus primarily inhibitory while at the same time avoiding some of the issues raised by the traditional conditioned inhibition paradigm, namely second-order conditioning, external inhibition, and configural learning. Using the fear-potentiated startle paradigm, we adapted an AX+, BX– training procedure in which stimuli A and X were presented simultaneously and paired with shock, and stimuli B and X were presented simultaneously in the absence of shock. In testing, high levels of fear-potentiated startle were seen in the presence of A and AX and much lower levels were seen in the presence of B and AB, as would be predicted if stimulus B were a conditioned inhibitor. We believe this method is a viable alternative to the traditional conditioned inhibition training procedure and will be useful for studying the neural mechanisms of fear inhibition.

A great deal is now known about the behavioral and neural characteristics of fear acquisition, thanks in large part to the study of Pavlovian fear conditioning. In this paradigm an animal is exposed to pairings of an initially neutral stimulus such as a light or tone (the conditioned stimulus or CS) with a mild footshock (the unconditioned stimulus or US), and comes to exhibit a fear-conditioned response (CR) in the presence of the CS. "Fear" is operationally defined in several ways, including freezing, ultrasonic vocalization, and an increase in the amplitude of an acoustic startle response, and is observable following a single CS-US pairing under some circumstances (Paschall and Davis 2002). Fear conditioning is thus an extremely robust form of learning and as a model system it has lent itself well to neural analyses on systems, cellular, and molecular levels (LeDoux 1996; Davis 2000).

Surprisingly, and in contrast to the extensive literature on fear acquisition, much less is known about the mechanisms of fear suppression or inhibition although this question is receiving increasing interest due to its clear clinical relevance (Bouton 2000). The reasons for this oversight are numerous, but primary among them is the lack of a satisfactory experimental paradigm for the study of inhibitory fear learning. Those studies that have addressed this issue almost always have focused on extinction, an experimental procedure in which a feared CS is presented repeatedly in the absence of the US, leading to a reduction or elimination of the fear CR (Pavlov 1927; Myers and Davis 2002). Extinction is attractive in many regards; in particular, it is convenient and sufficiently simple to be subjected to cellular and molecular analyses of the sort already applied to fear acquisition. However, it may also be argued that extinction, by itself, provides relatively limited information that must be accepted with a certain degree of caution.

Extinction has been studied for decades and much has been learned about its behavioral characteristics and potential neural underpinnings (Bouton 1993; Rescorla 2001). Among the most important insights to have emerged from this research is the understanding that extinction is a form of learning in its own right, rather than an "unlearning" or "forgetting" process that erases previous learning. In the associative language of psychological theories, behavioral extinction reflects the development of an inhibitory CS-US association that acts in parallel with an excitatory association (which itself was acquired through CS-US pairings) and directly opposes the tendency of that association to activate the US representation. Thus, following extinction, a CS is endowed with both excitatory and inhibitory tendencies; that is, it retains the capacity to generate a CR under some, but not all, circumstances (Konorski 1967; Bouton 1993).

However, this dual nature of the CS is problematic for neural analyses of inhibitory fear learning. For example, it is difficult to know whether a manipulation that retards extinction does so by impairing the development of inhibition or facilitating the expression of excitation, as either of these possibilities should, in principle, be reflected identically in behavior. The fact that many investigators apply permanent manipulations, such as electrolytic or chemical lesions, prior to acquisition of fear itself only compounds this interpretive problem (for discussion see Myers and Davis 2002). Ideally one would like a method of isolating the inhibitory component of extinction to allow an independent analysis of inhibition.

In fact, there are experimental procedures other than extinction that can endow a stimulus with an inhibitory association. The best known of these, called conditioned inhibition (CI) training, involves training one CS, A, as a conditioned excitor (i.e., A-US trials—often referred to as A+) and then presenting the

[4]Corresponding author.
E-MAIL mdavis4@emory.edu; FAX (404) 727-3233.

to-be-inhibitory stimulus, B, in compound with A and omitting the US (i.e., AB-no US trials—often referred to as AB−). As a function of this training, the conditioned inhibitor acquires the ability to reduce the magnitude of the CR produced by A or another excitatory CS (a standard summation test of inhibition; Rescorla 1969). The conditioned inhibitor is also slowed in its acquisition of a CR relative to an associatively neutral stimulus if it is later paired with the US (a retardation test of inhibition; Rescorla 1969).

Surprisingly, conditioned inhibition of freezing—the most popular fear index in neural studies—has been examined only rarely (e.g., Gewirtz et al. 1997; Vouimba et al. 2000). Several years ago we (Falls 1993; Falls and Davis 1997) adapted the basic conditioned inhibition design for use in the fear-potentiated startle paradigm, in which the amplitude of an acoustic startle response is increased when startle is elicited in the presence of a fear CS relative to when startle is elicited in the presence of a neutral CS or in the absence of explicit stimulation (Brown et al. 1951; Davis and Astrachan 1978). Initially, Falls (1993) used simultaneous presentation of the fear CS (a light) and the to-be-conditioned inhibitor (a low-frequency noise) but had to abandon this procedure because there was substantial external inhibition of responding to the light that could not be reduced by nonreinforced pre-exposure of the low-frequency noise. External inhibition refers to an unconditioned decrement in responding to an excitatory CS when a second stimulus (which is often neutral, but may be excitatory or inhibitory; Lovibond et al. 2000) is presented just before or at the same time as the excitatory CS (Pavlov 1927). Most investigators believe that external inhibition reflects the drawing of attention away from the excitatory CS (e.g., Pavlov 1927). Because this effect is unconditioned, it probably results from entirely different mechanisms than does conditioned inhibition, and yet external inhibition is a potential confound in any experiment that involves a summation test consisting of pairing a putative conditioned inhibitor with an excitatory CS.

To get around this problem, Falls and Davis (1997) used a serial procedure in which the offset of the white noise coincided with the onset of the light on compound trials (a "zero-trace" arrangement). They did find evidence of conditioned inhibition to the noise, as there was substantial startle potentiation in the presence of the light and significantly less startle potentiation in the presence of the noise-light compound. Furthermore, the inhibitory capacity of the noise transferred to a separately trained excitor (a tactile "fan" stimulus) whose onset coincided with the offset of the noise (Falls 1993). In addition, excitatory conditioning of the noise was retarded when shock was presented shortly after noise offset (i.e., at the time when the inhibitory effect of the noise should have been maximal based on this serial compound training procedure; Falls and Davis 1997).

However, the noise was not solely inhibitory, because startle elicited in its presence was significantly potentiated. This probably was caused by second-order conditioning, in which a neutral CS that is presented in compound with an excitatory CS itself becomes somewhat excitatory, presumably because the excitatory (first-order) CS acts as an unconditioned stimulus (Rescorla 1980). In a sense then, this paradigm raises the same interpretive issues as does extinction, in that it is difficult to isolate the inhibitory learning component from the excitatory component for separate neural analysis. In addition, the serial procedure tends to favor occasion setting, a capacity acquired by a CS to modulate responding to another CS by predicting the delivery or omission of reinforcement. This phenomenon occurs through different mechanisms than, and is independent of, the ability of the occasion setter to elicit or inhibit a response on its own (for review, see Swartzentruber 1995).

Thus, previous efforts to develop a paradigm for the study of inhibition of fear-potentiated startle have faced a choice among contamination by excitation as acquired through second-order conditioning, interference with expression of learned behavior via external inhibition, and occasion setting, which may involve mechanisms other than inhibition of fear. Clearly, each of these alternatives presents its own problems for a neural analysis of fear inhibition, and yet the benefits of adapting a procedure to the fear-potentiated startle paradigm that permits a less contaminated measure of fear inhibition are significant enough to warrant continuing effort in this domain. In this paper we examine external inhibition of fear-potentiated startle in more detail and assess various procedures for minimizing its contribution to response inhibition while maintaining a simultaneous compound stimulus presentation that largely circumvents the problems posed by occasion setting and, along with other modifications, reduces second-order conditioning. On the basis of these findings, we present a modified version of the conditioned inhibition design used by Falls and Davis (1997), which we believe is a superior method of analyzing inhibition in the fear-potentiated startle paradigm.

Experiment 1: External Inhibition of Fear-Potentiated Startle

External inhibition initially was described by Pavlov (1927) as an unconditioned suppression of the salivary CR upon presentation of an extraneous stimulus, such as a sudden change in illumination or an unexpected sound. Falls (1993) conducted a preliminary investigation of this phenomenon within the fear-potentiated startle paradigm and found that suppression of potentiation to a previously trained light was observable when the light co-occurred with a novel noise stimulus, consistent with Pavlov's observations. Interestingly, the magnitude of external inhibition was lessened considerably when the onset of the noise preceded the onset of the light by 3.7 sec (a zero-trace compound stimulus arrangement), suggesting that the temporal relation between the light and noise stimuli is related to the degree of external inhibition observed.

Experiment 1 was designed to explore more thoroughly the conditions under which external inhibition of fear-potentiated startle occurs. As in the Falls (1993) experiment, rats were fear-conditioned to a light CS and subsequently were tested for startle potentiation to the light, both when the light was presented in isolation and when it was compounded with a novel noise stimulus. The temporal placement of the noise with respect to the light varied across trials and included trace, zero-trace, simultaneous, and "backwards" arrangements, as illustrated in Figure 1. The experiment also included a "random" control group in which the light and shock generally were unpaired in training (see Materials and Methods for more details), which allowed us to determine whether the apparent external inhibition effect actually reflected unconditioned suppression of baseline startle by the novel noise stimulus.

MATERIALS AND METHODS

Animals

Sixteen male Sprague-Dawley rats (Charles River) weighing 350–450 g were used. Animals were maintained on a 12/12-h light-dark cycle (lights on at 0700) with food and water continuously available. All rats were housed in $45 \times 24 \times 20$-cm polycarbonate cages (four rats each) in a temperature-controlled (24°C) animal colony.

Apparatus

Animals were trained and tested in $8 \times 15 \times 15$-cm Plexiglas and wire-mesh cages. The cage floor consisted of four 6.0-mm diam-
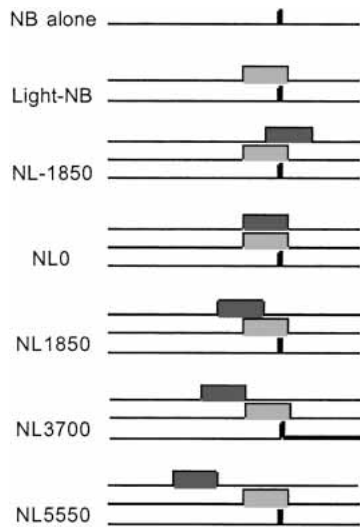
Figure 1 Schematic representation of the test trials of Experiment 1. Startle was assessed in each test trial via the presentation of a 95-dB, 50-msec white noise burst (black rectangles). In noise burst (NB)-alone trials, the startle stimulus was presented in isolation. In light-NB trials, the startle stimulus occurred 3.2 sec after the onset of a 3.7-sec light CS (light gray rectangles). Comparison of startle on light-NB and NB-alone trials provided an index of potentiated startle to the light. In five other trial types, the light and startle stimuli were presented as in the light-NB trials and were accompanied by a 3.7-sec white noise CS (dark gray rectangles). The temporal placement of the white noise with respect to the light differed among noise/light (NL) trials. In NL-1850 trials, the onset of the noise occurred 1850 msec after the onset of the light; in NL0 trials, the noise and light were presented simultaneously; in NL1850, NL3700, and NL5550 trials, the onset of the noise preceded the onset of the light by 1850, 3700, and 5550 msec, respectively. Light onsets occurred at a fixed 30-sec interstimulus interval.

eter stainless steel bars spaced 18 mm apart. Each cage was suspended between compression springs within a steel frame and located within a custom-designed 90 × 70 × 70-cm ventilated sound-attenuating chamber. Background noise (60 dB wideband) was provided by a General Radio Type 1390-B noise generator and delivered through high-frequency speakers (Radio Shack Supertweeter) located 5 cm in front of each cage. Sound level measurements (sound pressure level) were made with a Bruel and Kjaer model 2235 sound-level meter (A scale; random input) with the microphone (Type 4176) located 7 cm from the center of the speaker (approximating the distance of the rat's ear from the speaker). A red light bulb (7.5 W) located 25 cm from the stabilimeter illuminated the chamber at all times.

Startle responses were evoked by 50-msec, 95-dB white noise bursts (5 msec rise-decay) generated by a Macintosh G3 computer soundfile (0–22 kHz), amplified by a Radio Shack amplifier (100 W; model MPA-200; Tandy), and delivered through the same speakers used to provide background noise. An accelerometer (model U321A02; PCB Piezotronics) affixed to the bottom of each cage produced a voltage output proportional to the velocity of cage movement. This output was amplified and rectified to give cage velocity as the output (model 483B21; PCB Piezotronics) and digitized on a scale of 0–2500 arbitrary units by an InstruNET device (model 100B; GW Instruments) interfaced to a Macintosh G3 computer. Startle amplitude was defined as the maximal peak-to-peak voltage that occurred during the first 200 msec after onset of the startle-eliciting stimulus.

A 3.7-sec light CS (80 lux) was produced by an 8 W fluorescent bulb (100 µsec rise time) located 10 cm behind each cage. Luminosity was measured using a VWR light meter. The noise CS was a 3.7-sec, 75-dB (SPL, A-scale) white noise that was band-pass

filtered with high and low passes both set at 2 kHz (Campeau and Davis 1992). The noise was delivered through a 90-watt 3-way speaker (Radio Shack #12-1767) mounted to the rear interior face of the sound-attenuating chamber. The US was a 0.5-sec shock, delivered to the floorbars and produced by a shock generator (SGS-004; LeHigh Valley). Shock intensities (measured as in Cassella and Davis 1986) were 0.4 mA.

The presentation and sequencing of all stimuli were under the control of the Macintosh G3 computer using custom-designed software (The Experimenter; Glassbeads).

Procedure

Matching

On each of two days, rats were placed in the startle chambers and 5 min later presented with 30 startle stimuli (95 dB; 30-sec interstimulus interval—ISI). The rats subsequently were matched into two groups of eight rats each, with each group having a similar mean startle amplitude based on the 30 startle stimuli of the second matching session.

Training

Twenty-four h after the second matching session, rats were returned to the startle chambers for the first of two daily training sessions. Five minutes after being placed in the chambers, rats assigned to the paired group received the first of 10 presentations of a 3.7-sec light (conditioned stimulus—CS) that coterminated with a 0.5-sec, 0.4-mA footshock. The mean intertrial interval (ITI; for training, defined as the interval between the onsets of successive CSs) was 4 min (range, 3–5 min). Rats assigned to the "random" group received footshocks at the same points in time as for the paired group; however, the lights were scheduled to occur "randomly" in time so as to minimize any light-footshock contingency. Specifically, the training session was divided into 5-sec "bins," 10 of which were occupied by footshocks. The 10 lights were assigned randomly to occupied or unoccupied bins, such that light-footshock pairings were possible but unlikely. (One of 10 lights was paired with a footshock in our sequence.) Five min after the final training event, rats were removed from the chambers and returned to their home cages.

Testing

Twenty-four h after the second training session, rats were returned to the startle chambers and, 5 min later, received the first of 30 95-dB noise bursts (30-sec ISI). These initial startle stimuli were meant to produce a stable startle reflex baseline before introduction of the CSs, and were followed immediately by a test of startle potentiation to the light and to compounds of the light and a novel white noise stimulus (75 dB, 3.7 sec). In all, there were seven trial types: noise burst (NB) alone, consisting of presentations of the startle stimulus in the absence of the light CS; light-NB, consisting of presentations of the startle stimulus 3.2 sec after the onset of the light; and trials similar to the light-NB trials except that the white noise was presented just before, during, or after the onset of the light. There were five different noise/light-NB (NL) trial types, referred to as NL-1850, NL0, NL1850, NL3700, and NL5550 (see Fig. 1). In NL0 trials, the onset of the noise occurred simultaneously with the onset of the light; in NL1850, NL3700, and NL5550 trials, the onset of the noise preceded the onset of the light by 1850, 3700, and 5500 msec, respectively; and in NL-1850 trials the onset of noise occurred 1850 msec after the onset of the light. The interval between startle stimuli was fixed at 30 sec. There were 10 blocks of seven trials, and the trial types were pseudorandomly arranged with the restriction that each trial type occur once within each trial block.

Statistical Analysis

Mean startle amplitudes across the 10 occurrences of each trial type were computed for each rat. Difference scores were calculated for each rat by subtracting the mean startle amplitude obtained on NB-alone test trials from the mean startle amplitude on light-NB and noise/light-NB trials. The resulting difference scores reflect the magnitude of fear-potentiated startle in the presence of the light or noise/light compound. The data were analyzed with a mixed-model ANOVA with group as a between-subjects factor and trial type as a repeated measure. Post-hoc analyses included lower-order ANOVAs and tests for quadratic trends.

RESULTS AND DISCUSSION

The data from the test session are presented in Figure 2. The top panel shows the mean difference scores on the light-NB and NL trial types in the paired and "random" groups. As expected, the rats of the random group did not exhibit measurable startle potentiation in the presence of the light. In the paired group, in contrast, there was significant potentiation to the light and significant external inhibition on trials in which the noise was presented. External inhibition was maximal when the light and noise were presented simultaneously (i.e., NL0 trials), and was substantial on NL-1850 and NL1850 trials.

Statistical analyses confirmed these observations. There was a significant main effect of trial type [$F(5,70) = 12.96$, $P < 0.01$] and trial type × group interaction [$F(5,70) = 9.74$; $P < 0.01$], the latter of which showed a reliable quadratic trend [$F(1,14) = 20.74$; $P < 0.01$]. Analysis of the performance of each group separately via repeated-measures ANOVAs revealed a significant main effect of trial type [$F(5,35) = 13.46$; $P < 0.01$] and a significant quadratic trend [$F(1,7) = 27.57$; $P < 0.01$] in the paired group but not in the "random" group.
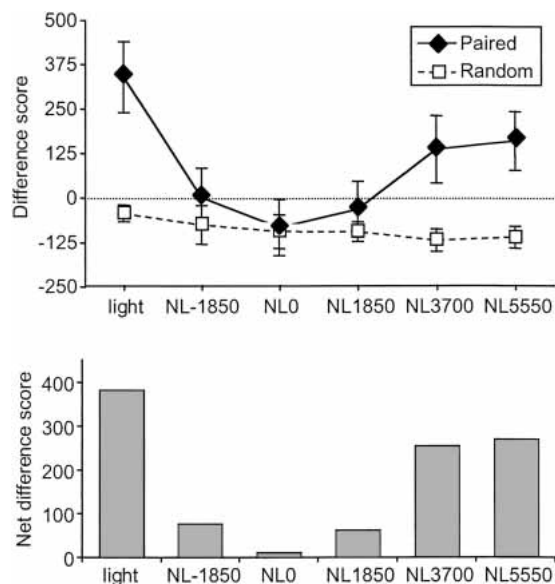

Figure 2 *Top:* Mean startle difference scores obtained on light-NB, NL-1850, NL0, NL1850, NL3700, and NL5550 trial types in the Paired and Random groups of Experiment 1. The dashed line at zero indicates no change in startle from baseline. Error bars represent 1 standard error of the mean (SEM). *Bottom:* The same data as in the top panel are shown, now transformed by subtracting the mean difference score of the Random group on each trial type from the mean difference score of the Paired group on the corresponding trial type. This transformation was intended to factor out any unconditioned effects of the light and noise on baseline startle so as to visualize the effects of these stimuli on potentiated startle.

To assess the effects of the light and noise on conditioned startle potentiation independently of any nonassociative effects on baseline startle, we subtracted the mean difference scores of the random group from those of the paired group (Davis et al. 1989). This was appropriate because there was no difference in startle amplitude on the NB-alone test trials between the two groups [$t(14) = .885$; $P > 0.05$]. These transformed data are presented in the lower panel of Figure 2. Again, it is evident that startle potentiation observed on NL trials was significantly less than that on light-NB trials, indicating external inhibition by the noise, and that the effect was strongest on NL0 trials.

These data replicate the finding of Falls (1993) that external inhibition of responding to a previously trained light by a novel noise stimulus is robust within the fear-potentiated startle paradigm, particularly when the onset of the noise coincides with the onset of the light. Moreover, systematic variation of the temporal placement of the noise with respect to the light produced a similar, orderly variation in the magnitude of external inhibition, indicating that the suppressive ability of the noise is time-limited and decays following noise offset. Importantly, comparison of the paired and random groups indicates that this effect is on startle potentiation (i.e., fear), rather than baseline startle.

The finding of substantial external inhibition with simultaneous compounds would seem to preclude their use within a conditioned inhibition training procedure that incorporates a summation test. However, because this stimulus arrangement largely circumvents the problem of occasion setting encountered by Falls and Davis (1997), we next sought to determine whether external inhibition observed with simultaneous compounds could be lessened in magnitude through one or more parametric variations. As mentioned earlier, Falls (1993) found that extensive pre-exposure to this 2-kHz noise only slightly reduced external inhibition. Hence, in Experiment 2 we used a less complex stimulus (a pure tone) as the external inhibitor, which we expected might be more sensitive to pre-exposure given the general finding that the rate of habituation is greater, the less complex the stimulus (e.g., Richardson et al. 1994).

Experiment 2: Factors Affecting the Magnitude of External Inhibition

Because he considered external inhibition to be an attentional phenomenon, Pavlov (1927) believed that repeated, nonreinforced exposure to an extraneous stimulus would lead to habituation of the orienting reflex to that stimulus and detract from its ability to suppress responding to a CS (p. 46). Consistent with this idea, Reiss and Wagner (1972) demonstrated that external inhibition of the conditioned eyeblink response in rabbits was much less robust when the extraneous stimulus had been extensively pre-exposed (1380 nonreinforced presentations) than when it had been presented relatively few times (12 nonreinforced presentations). One of the aims of Experiment 2 was to examine the pre-exposure effect within the fear-potentiated startle paradigm. Presumably, if the suppressive effect of the noise is due to its drawing attention away from the light, then pre-exposure to the noise prior to test should reduce its novelty and, by the same token, its salience.

In Experiment 1, the noise per se was not the only novel aspect of the test; compound stimuli also were experienced for the first time, and this may have contributed to an orienting reflex or general drawing of attention away from the light. If so, it is possible that exposure to stimulus compounds prior to test might reduce their novelty in much the same way that pre-exposure to the external inhibitor reduced its salience in the study of Reiss and Wagner (1972). Moreover, because the compound stimulus in the test is comprised of the excitor and exter-

nal inhibitor, then prior experience with each of these stimuli as part of a compound presumably would minimize the novelty of their being compounded with one another. Thus, training of the excitor in compound with another stimulus (say, a tone), and pre-exposure of the external inhibitor in compound with a fourth stimulus (say, a quiet fan), should be beneficial in the sense that it provides the animals with exposure to stimulus compounds, in general, and experience with the excitor and external inhibitor co-occurring with another stimulus, in particular. The second aim of Experiment 2 was to examine the effect of compound stimulus pre-exposure on the magnitude of external inhibition in test.

The experiment involved six groups of rats and four discriminable stimuli (light, noise, tone, and fan; Falls and Davis 1994). The stimulus assignments were somewhat different from those of Experiment 1, however, in that the light was the excitor and the tone was the external inhibitor. This choice was motivated by the finding of Falls (1993) that extensive pre-exposure to the 2-KHz noise only slightly reduced external inhibition. Because habituation is greater with less complex stimuli (e.g., Richardson et al. 1994), it seemed possible that a pure tone would be more sensitive to pre-exposure. The groups, which differed from one another in the conditions of their training, are described in Table 1.

Half of the rats received reinforced training with the excitor alone (the light, designated A) and half with the excitor in compound with another stimulus (the noise, designated X). All groups also received nonreinforced noise presentations (X−), which were intended to extinguish any excitation to X in the AX+ groups, and thereby allow A to acquire additional excitation (Rescorla and Wagner 1972). Theoretically these trials should have no effect on A's excitation in the A+ groups, and were included in their training solely for the purpose of equating stimulus exposure across groups. Through comparison of the A+ and AX+ groups, it is possible to determine the effect of presenting the excitor as part of a compound in training upon the magnitude of external inhibition in test.

The groups were further subdivided into three pairs, each comprising one A+ and one AX+ group, that were distinguished by their pre-exposure, or lack thereof, to the external inhibitor, a pure tone. One pair received no tone exposure; these groups were labeled A+/X−/0 and AX+/X−/0 (where "0" indicates the absence of any stimulus). A second pair of groups, labeled A+/X−/B− and AX+/X−/B−, received nonreinforced presentations of the tone (B). The remaining pair of groups, labeled A+/X−/BY− and AX+/X−/BY−, received nonreinforced presentations of the tone in compound with a tactile fan stimulus (Y). Through comparison of these three sets of groups, it is possible to determine whether experience with the external inhibitor, either in isola-

tion or as part of a compound, reduces the ability of that stimulus to suppress responding to the excitor in test. As in Experiment 1, all compounded stimuli were presented simultaneously.

MATERIALS AND METHODS

Animals

Sixty-eight male Sprague-Dawley rats (Charles River) weighing 350–450 g were used. The rats were housed and maintained as described in Experiment 1. The experiment was run in two replications involving 36 and 32 rats, respectively.

Apparatus

The apparatus was identical to that described in Experiment 1. In addition, a 3.7-sec, 75-dB tone CS (50-mec rise/decay time) was band-pass filtered with high and low passes both set at 2 kHz, and was presented via the same speaker as the white noise CS. A 3.7-sec tactile "fan" CS was provided by 12-V DC, 3-in brushless computer fans (Radio Shack; model #273-243) mounted to the top exterior portion of each startle cage (Falls and Davis 1994). Airflow was through a hole in the top of the Plexiglas cage and was vertical and downward. Sound pressure measurements made with a sensitive Bruel and Kjaer decibel meter (model 2235) revealed that the fan CS did not raise the overall noise level above the background.

Procedure

Matching

Matching proceeded as described in Experiment 1. Rats were assigned to six groups (8–10 rats per group) exhibiting equivalent mean startle amplitudes.

Training

Twenty-four h after the second matching session, rats were returned to the startle chambers for a single training session. Five min after being placed in the chambers, the rats received the first of 30 training trials. Table 1 provides an outline of the trial types presented to each of the six groups. In all groups, A was the light, X was the white-noise CS, B was the tone, and Y was the fan. Compound stimuli were presented simultaneously; that is, the onsets of compounded CSs co-occurred. The intertrial interval (ITI) was fixed at 60 sec. In groups A+/X−/0 and AX+/X−/0, "dummy" trials in which no stimuli were presented (denoted "0") occurred at the same points in time as did the B− and BY− trials for the other groups. There were 10 blocks of three trials, each comprising one presentation each of the stimulus types appropriate to each group (i.e., one presentation of A+ or AX+; one presentation of X−; and one presentation of B−, BY−, or 0). The trial types were pseudorandomly arranged. Five min after the final training event, rats were removed from the chambers and returned to their home cages.

Testing

Twenty-four hours later, rats were returned to the startle chambers for a test of startle potentiation to the light (A) and to a simultaneous compound of the light and the tone (AB). Five min after being placed in the chambers, the rats received the first of 30 95-dB noise bursts (30-sec ISI) intended to produce a stable startle baseline. These were followed immediately by 40 additional startle stimuli, of which 10 occurred 3.2 sec after the onset of the light, 10 occurred 3.2 sec after the onset of the light/tone compound, and 20 occurred in the absence of other stimuli. There were 10 blocks of four trials, each including one light-NB, one light/tone-NB, and two NB-alone trials. Within each block, the trial types were pseudorandomly arranged.

Table 1. Training Events in Experiment 2

Group	Training trial types		
	Excitor training	X extinction	External inhibitor pre-exposure
A+/X−/0	A+	X−	0
AX+/X−/0	AX+	X−	0
A+/X−/B−	A+	X−	B−
AX+/X−/B−	AX+	X−	B−
A+/X−/BY−	A+	X−	BY−
AX+/X−/BY−	AX+	X−	BY−

Note. "+" denotes shock reinforcement; "−" denotes omisssion of reinforcement; "0" denotes no stimulus; and AX denotes a simultaneous compound in which the onset of A is coincident with the onset of X.

Rats exhibiting less than 25% potentiated startle to the light in test, defined as [(startle amplitude on light-NB minus NB alone trials)/NB alone trials] × 100, were discarded. This is because external inhibition of potentiation to the light can be accurately assessed only when there is a reasonable level of potentiation to inhibit. Sixteen rats (23.5%) were eliminated from the data analysis on this basis.

Statistical Analysis

Difference scores were calculated for each rat by subtracting the mean startle amplitude obtained on NB-alone trials from the mean startle amplitude on light-NB and light/tone-NB trials. The data were analyzed with a mixed-model ANOVA, with excitor training (A+, AX+) and external inhibitor pre-exposure (none, B, BY) as between-subjects factors and test trial type (A, AB) as a repeated measure. Follow-up analyses included lower-order ANOVAs, post-hoc tests for linear trends, and Bonferonni tests.

RESULTS AND DISCUSSION

The data from the test are presented in Figure 3. It is evident that, in the two leftmost groups, which had no prior exposure to the tone (B), simultaneous presentation of B in compound with the light (A) during the test session produced strong external inhibition. This is consistent with the marked external inhibition observed on NL0 trials in Experiment 1 using the novel noise stimulus. The magnitude of external inhibition did not differ between groups that were trained with A+ or AX+, indicating that experience with the excitor (A) in compound with another stimulus (X) did not reduce the magnitude of external inhibition when A was subsequently paired with a novel stimulus (B). In contrast, prior exposure to B (middle groups) or BY (rightmost groups) did markedly reduce external inhibition on AB test trials.

Statistical analyses supported these observations. The interaction between test cue (A, AB) and external inhibitor pre-exposure training (none, B−, BY−) was reliable [$F_{(2,46)} = 5.81$; $P < 0.01$] and showed a significant linear trend [$F_{(2,46)} = 5.81$; $P < 0.01$], indicating that the magnitude of external inhibition by B was greatest in the groups that had no pre-exposure to B, moderate in the groups that had pre-exposure to B by itself, and least in the groups that had pre-exposure to B as part of a BY compound. The interaction between test cue (A, AB) and excitor training condition (A+, AX+) just missed significance [$F_{(1,46)} = 3.37$; $P = 0.07$], suggesting a modest reduction in the magnitude of external inhibition to A when A was trained as part of an AX compound, relative to when A was trained by itself.
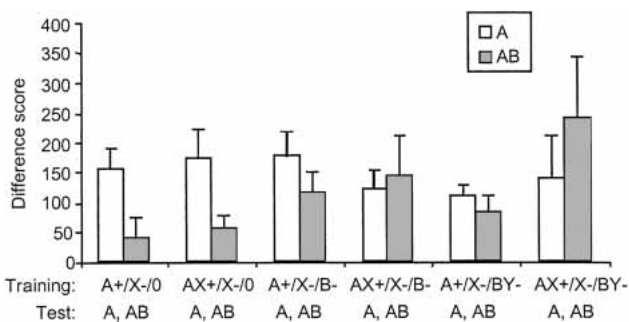

Figure 3 Mean startle difference scores obtained on the A and AB test trials of Experiment 2, in the six groups of rats specified in Table 1. Error bars, 1 SEM.

Although the three-way interaction of test cue (A, AB), excitor training (A+, AX+), and external inhibitor pre-exposure (none, B−, BY−) did not quite reach significance, it does appear from the figure that there was an additional reduction in the magnitude of external inhibition when excitor compound training (AX+) was combined with pre-exposure to the external inhibitor as part of a compound (BY−). This observation is supported by a significant main effect of external inhibitor pre-exposure in the groups trained with AX+ [$F_{(2,22)} = 4.54$; $P < 0.05$] but not in the groups trained with A+ [$F_{(2,24)} = 1.31$; $P > 0.05$] when the difference in responding to A and AB is taken as a measure of the magnitude of external inhibition. A post-hoc Bonferroni test performed consequent to the main effect of external inhibitor pre-exposure in the AX+ groups indicated that only the difference between group A+/X−/0 and AX+/X−/BY− was reliable [$P < 0.05$].

These findings indicate that although external inhibition of potentiation to a previously trained light stimulus is robust when the light is presented simultaneously with a novel tone stimulus, this effect may be minimized or even eliminated altogether with a combination of compound excitor training and compound external inhibitor pre-exposure. This is consistent with an attentional interpretation of external inhibition, because experience with the external inhibitor as well as with stimulus compounds presumably minimizes their novelty in test when the light is paired for the first time with the tone. With this information in hand, we can return to the question of most interest, namely whether it is possible to devise a conditioned inhibition training procedure for the fear-potentiated startle paradigm that circumvents the problems posed by occasion setting, second-order conditioning, and external inhibition.

Experiment 3: Development of a Conditioned Inhibition Training Procedure

The standard protocol for training conditioned inhibition, as we have seen, assumes the form A+, BA−, where A is reinforced when presented in isolation and nonreinforced when accompanied by B, the conditioned inhibitor. In their studies of inhibitory fear learning with this procedure, Falls and Davis (1997) presented the BA compound in a zero-trace arrangement, contrary to the simultaneous stimulus presentation favored by most investigators (Holland 1985). Their primary reason for doing so was to minimize the potential for external inhibition of responding to A by B, but Experiment 2 suggests that external inhibition may have been less of a problem than they anticipated when using a tone as the to-be-conditioned inhibitor. That is, because their presumed external inhibitor, B, was pre-exposed in training, and because the animals had experience with compound stimuli prior to test, external inhibition likely would have been small even if their AB test compound was simultaneous rather than zero-trace. However, Experiment 2 also demonstrates that external inhibition may be minimized to the greatest extent, and perhaps even eliminated altogether, with a combination of compound excitor training and compound external inhibitor pre-exposure, as in group AX+/X−/BY−. If it were possible to incorporate these features into a conditioned inhibition training protocol, the resulting design could have certain advantages over the traditional CI design, at least within the fear-potentiated startle paradigm.

In fact, the A+, BA− discrimination is only one of several conditioned inhibition training procedures. Theoretically, a stimulus will acquire inhibition whenever its occurrence is correlated with the omission of a predicted or "expected" US, a condition that may be met in several ways. More formally speaking, a CS becomes a conditioned inhibitor when it is nonrein-

forced against a background of excitation (i.e., an expectation of US delivery), which most commonly is provided by a concurrently presented, previously reinforced CS (Wagner and Rescorla 1972). Thus, if we were to begin with the AX+/X−/BY− group of Experiment 2, we could modify its training conditions so as to generate a conditioned inhibitor while, at the same time, maintaining the dual compound arrangement that is so favorable in terms of minimizing external inhibition. The most straightforward way of doing so would be to eliminate the X− trials and substitute BX− trials for BY− exposure, an AX+, BX− design. Here, X should retain the excitation it acquires on AX+ trials because it is not extinguished via X− trials, and B should become a conditioned inhibitor as a function of its being nonreinforced in the presence of X. As in the standard conditioned inhibition training protocol, B's inhibitory capacity can be examined following training on the discrimination by comparing startle potentiation in the presence of A, which should be large, and AB, which should be relatively small if indeed B has acquired inhibition.

It is possible to generate a more formal prediction as to the outcome of the AX+, BX− discrimination by applying a quantitative theory of Pavlovian conditioning, such as the Rescorla-Wagner model (1972), to the problem. The learning rule of the model is represented mathematically as follows:

$$\Delta V_i = \alpha\beta \, (\lambda \, - \, \Sigma V_i),$$

where V_i denotes the predictiveness of reinforcement of CS_i (and V_i indicates the change in that predictiveness occurring on any single trial); α and β are learning rate parameters associated with the CS and US, respectively; and λ represents the US, assuming a value of 1 when the US occurs and 0 when the US is omitted. When applied to an alternating sequence of AX+ and BX− trials, the Rescorla-Wagner model predicts that the discrimination will proceed along the course described in Figure 4: There will be an initial increase in responding to both AX and BX, followed by a decline in responding to BX and continuing growth in responding to AX until asymptote is reached. The predicted responding on AX, BX, AB, A, B, and X postdiscrimination test trials, which include the critical summation test of responding to A and AB, are also presented in Figure 4, and indicate that the discrimination is solved when A becomes strongly excitatory, X becomes moderately excitatory, and B becomes as inhibitory as X is excitatory.
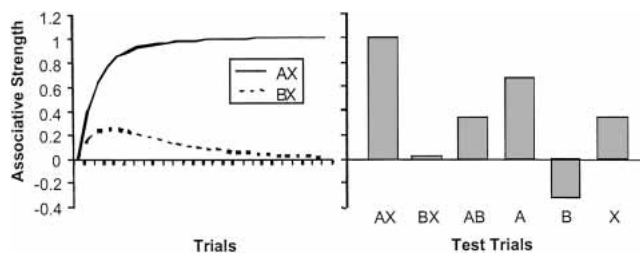

Figure 4 *Left:* Changes in the net associative strength of the AX+ and BX− stimuli of Experiment 3, as predicted by a computer simulation of the Rescorla-Wagner model (1972). Initially animals are predicted to respond to both AX and BX, but as the discrimination progresses, responding to BX returns to zero while responding to AX increases until asymptote is reached. Parameter values were set as follows: 0.2 on all trial types; 1 on reinforced trials and 0.5 on nonreinforced trials; 1 on reinforced trials and 0 on nonreinforced trials. *Right:* Predicted associative strength of A, B, and X, as well as AX, BX, and AB compounds, when the discrimination has reached asymptote. A is predicted to become strongly excitatory, X to become moderately excitatory, and B to become as inhibitory as X is excitatory. Comparison of responding on A and AB test trials reveals the inhibition that has accrued to B.

The AX+, BX− discrimination has been examined before, under different circumstances and for different reasons, by Wagner et al. (1968). Those investigators were most interested in the differential associative strength accruing to the A, B, and X stimuli in separate groups of animals trained with AX+, BX−, and AX, BX (the latter of which denotes a pseudodiscrimination procedure in which AX and BX are reinforced on 50% of their presentations). Consistent with a contingency-based account of associative learning, A became strongly excitatory, and B became moderately inhibitory, with AX+, BX− training, whereas both A and B were moderately excitatory following AX, BX training. Responding to X was moderate and approximately equal in the two groups. Wagner and Rescorla (1972) subsequently applied their model to these two conditions, and the results reported in Figure 4 for the AX+, BX− condition are a replication of their findings.

Relatively complex discrimination procedures such as this one have been examined only rarely in fear conditioning, and only then when freezing was taken as the measure of fear (Rickert et al. 1979). It is not at all clear, therefore, whether the complex pattern of responsiveness to the A, B, X, and compound cues as predicted by Wagner and Rescorla will be evident in startle modulation. Experiment 3 was designed to evaluate this issue. Rats were exposed to three daily sessions of AX+, BX− training, where A, B, and X were light, noise, and fan stimuli whose assignment was counterbalanced across rats. Twenty-four h after the completion of training, all rats were tested for startle potentiation in the presence of A, B, X, AX, BX, and AB.

MATERIALS AND METHODS

Animals
Twenty male Sprague-Dawley rats (Charles River) weighing 350–450 g were used. The rats were housed and maintained as described in Experiment 1.

Apparatus
The apparatus was identical to that described in Experiment 2.

Procedure

Matching
Matching proceeded as described in Experiment 1. Rats were assigned to six groups (3–4 rats per group) exhibiting equivalent mean startle amplitudes.

Pretest
Twenty-four h after the second matching session, rats were returned to the startle chambers for a test of startle in the presence of the light, noise, fan, and compounds of these stimuli. Five min after being placed in the chambers, the rats received the first of 30 95-dB noise bursts (30-sec ISI) intended to produce a stable startle reflex baseline. These were followed immediately by 40 additional startle stimuli, of which 10 occurred in the absence of other stimuli (NB-alone trials) and 30 occurred 3.2 sec after the onset of the light, noise, fan, light/noise, light/fan, and fan/noise stimuli. There were five blocks of seven trials, each including two NB-alone trials and one each of A, B, X, AX, BX, and AB. The trial types were pseudorandomly arranged. Compound stimuli were presented simultaneously.

Training
Twenty-four h after the pretest, rats were returned to the startle chambers for the first of three daily training sessions. Five min after being placed into the chambers, the rats received the first of 20 training trials, of which half were AX+ and half were BX−
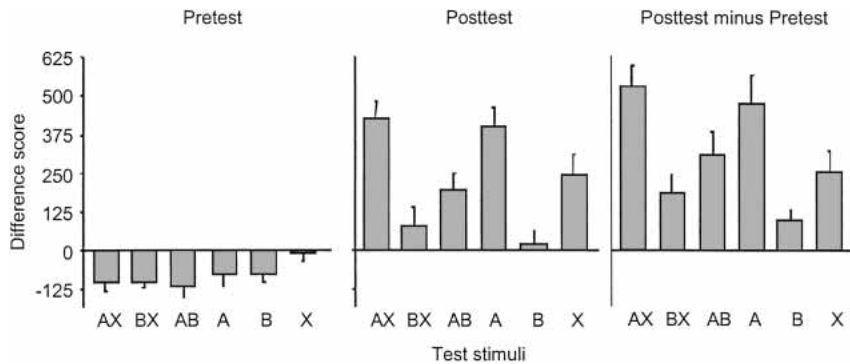
Figure 5 Mean startle difference scores obtained on AX, BX, AB, A, B, and X trial types in the pretest (*left* panel) and posttest (*center* panel) of Experiment 3. The *right* panel presents the same data, transformed by subtracting the mean difference score of each rat on each trial type in the pretest from its corresponding mean difference score in the posttest. Error bars, 1 SEM.

(3-min variable ITI; range 3–5 min). The trial types were pseudo-randomly arranged with the condition that no one trial type occurred more than twice in succession. Assignment of light, noise, and fan cues as A, B, and X was fully counterbalanced, such that each group of rats received a different cue assignment. The light, noise, and fan stimuli were 3.7 sec in duration, and the 0.5-sec footshock occurred 3.2 sec after the onset of AX. Compound stimuli were presented simultaneously. Five min after the final training event, rats were removed from the chambers and returned to their home cages.

Posttest

Twenty-four h after the final training session, rats were returned to the startle chambers for a test of startle potentiation to the light, noise, fan, light/noise, light/fan, and fan/noise stimuli. The posttest was identical to the pretest except that there were twice as many test trials (i.e., 20 NB-alone trials and 10 each of A, B, X, AX, BX, and AB).

Statistical Analysis

Mean startle amplitude across the 10 occurrences of each trial type were computed for each rat. Difference scores were calculated for each rat in each test by subtracting the mean startle amplitude obtained on NB-alone trials from the mean startle amplitude obtained on A, B, X, AB, AX, and BX trials, respectively. The data were analyzed with repeated-measures ANOVAs with trial type as a factor. Comparisons of difference scores on the various trial types were made with lower-level ANOVAs, using the error term from the overall ANOVA to control familywise error.

RESULTS AND DISCUSSION

The data from the pretest and posttest are presented in the leftmost and center panels, respectively, of Figure 5. It is evident from the pretest data that neither A, B, nor X, nor any compound thereof, produced any startle potentiation prior to training. There was some depression of startle with each stimulus or stimulus compound, an effect that appeared to be least pronounced on the X trials. This apparent difference is of questionable significance, however, because the assignment of light, noise, and fan cues as A, B, and X was counterbalanced across rats. In the posttest, in contrast, there were marked differences in the magnitude of startle potentiation in the presence of the various test stimuli. First, the rats did discriminate between AX and BX, as indicated by the greater potentiation in the presence of AX than BX. Second, comparison of AB and A indicates that B was inhibitory, as

potentiation to AB was less pronounced than potentiation to A. Finally, potentiation to B was minimal, whereas potentiation to X was measurable but less than that to A, consistent with the prediction that B should acquire inhibition by virtue of its being nonreinforced in the presence of X.

Statistical analyses supported these observations. In the pretest there was a significant main effect of trial type [$F(5,95) = 13.17$; $P < 0.01$], which was due entirely to the relatively small difference score on X trials. Thus, lower-level ANOVAs using the error term from the overall analysis revealed significant differences between X and every other trial type [F's > 5.30; P's < 0.05] and no other differences between any other pair of trial types. In the posttest there was, likewise, a reliable main effect of trial type [$F(5,95) = 11.41$; $P < 0.01$]. A significant difference was observed between AX and BX, indicating reliable discrimination [$F(1,95) = 25.14$, $P < 0.01$], and between AB and A, indicating conditioned inhibition by B [$F(1,95) = 8.54$; $P < 0.01$]. Reliable differences were also observed between AX and AB, AB and B, A and B, A and X, and B and X [F's > 5.07; P's < 0.05].

We also examined responding to the A and AB test stimuli across the duration of the session to determine whether B exerted inhibition from its first presentation or, alternatively, whether its suppression of responding to A was evident only after several AB presentations. The data are presented in Figure 6, which plots mean startle amplitude on NB-alone, A-NB, and AB-NB trial types. It is clear from the figure that there was a strong difference in responding to A and AB, and that this difference was apparent from the first presentation of these stimuli through the duration of the test session. A paired-samples t-test comparing responding on the first presentations of A and AB indicated that this difference was significant [$t(19) = 2.77$; $P < 0.05$].

Because there was a significant, unconditioned depression of startle in the pretest, we sought to assess the effects of the various cues on startle potentiation independently of any nonassociative effects on baseline startle. To do so we subtracted the mean difference score of each rat, on each trial type, in the pretest from its corresponding mean difference score in the posttest. This was appropriate because there was no difference between the two tests in startle amplitude on the NB-alone test trials [$t(19) = .276$; $P > 0.05$]. The transformed data are presented in the rightmost panel of Figure 6, which is very similar to the center panel in the pattern of responsiveness to the various cues that it presents. Statistical analyses of these data were identical to those of the posttest in the effects and differences identified as significant.

The pattern of responsiveness across the various test trial types is remarkably similar to that predicted by the Rescorla-Wagner model (1972; Fig. 4), suggesting that the discrimination is solvable by rats in this paradigm and that the mechanisms by which it is solved are well described by the model. Of most import is responding to A, B, and AB, because these trials provide good evidence that B was inhibitory at the conclusion of the discrimination. Thus, even though there was some minimal startle potentiation in the presence of B, comparison of A and AB indicates that B nevertheless was sufficiently inhibitory to suppress responding controlled by A. Given the outcome of Experiment 2, there is little reason to believe that this difference is due to an attentional external inhibition effect as opposed to a conditioned inhibitory learning process.
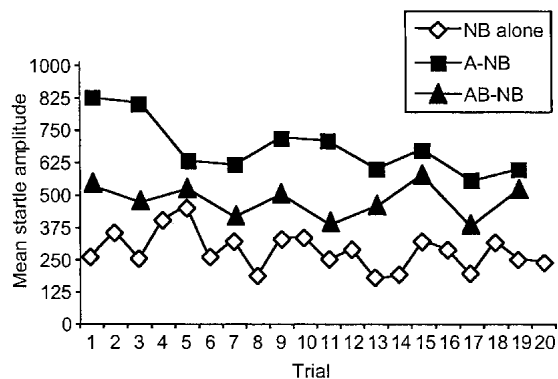
Figure 6 Mean startle amplitude on the individual A and AB test trials of Experiment 3. Startle was elevated on both A and AB test trials relative to NB-alone trials, but the magnitude of this difference was much greater on A trials than on AB test trials, consistent with an inhibitory effect of B on responding to A. Importantly, the difference in responding between A and AB was evident from the first presentations of these stimuli in test, and remained robust throughout the session.

Experiment 4: Acquisition of the AX+, BX– Discrimination

Experiment 3 indicates that rats can learn the AX+, BX– discrimination with sufficient exposure to AX and BX, but says nothing about the course of the discrimination or the number of trials required for differential responding to emerge. Experiment 4 was designed to explore these issues, and involved three separate groups of rats that received one, two, or three daily sessions of AX+, BX– training. Each group was tested for responding to A, B, X, AX, BX, and AB 24 h after the completion of its training regimen. As in Experiment 3, assignment of light, noise, and fan stimuli as A, B, and X was fully counterbalanced across rats.

MATERIALS AND METHODS

Animals

Forty male Sprague-Dawley rats (Charles River) weighing 350–450 g were used. The rats were housed and maintained as described in Experiment 1.

Apparatus

The apparatus was identical to that described in Experiment 2.

Procedure

Matching

Matching proceeded as described in Experiment 1. Rats were assigned to three groups (13–14 rats per group) exhibiting equivalent mean startle amplitudes.

Pretest

The pretest occurred 48 h after the second matching session, and proceeded as described in Experiment 3.

Training

Forty-eight h after the pretest, rats were returned to the startle chambers. Separate groups of rats received one, two, or three training sessions separated by 48-h intervals. Five min after being placed in the chambers, the rats received the first of 10 training trials, of which half were AX+ trials and half were BX– trials. Assignment of light, noise, and fan cues as A, B, and X was fully counterbalanced across rats within each group. The light, noise, and fan stimuli were 3.7 sec in duration, and the 0.5-sec foot-shock occurred 3.2 sec after the onset of the appropriate com-

pound stimulus. Compound stimuli were presented simultaneously. The ITI was 60 sec. Five min after the final training event, rats were removed from the chambers and returned to their home cages.

Posttest

Twenty-four h after each group's final training session, rats of that group were returned to the startle chambers for a test of startle potentiation to the light, noise, fan, light/noise, light/fan, and fan/noise stimuli. The posttest proceeded as described in Experiment 3.

Statistical Analysis

Difference scores were calculated for each rat by subtracting the mean startle amplitude obtained on NB-alone trials from the mean startle amplitude obtained on A, B, X, AB, AX, and BX trials, respectively. The data were analyzed with a mixed-model ANOVA with test day as a between-subjects factor and test cue as a repeated measure. Follow-up analyses included lower-order ANOVAs using the error term from the overall ANOVA to control familywise error.

RESULTS AND DISCUSSION

The course of the discrimination is presented in the top panel of Figure 7, which indicates that there was no potentiation to either AX or BX in the pretest, consistent with the pretest results from Experiment 3. Potentiation to AX did increase relative to pretest as a function of days of training, whereas potentiation to BX increased relative to pretest on test days 1 and 2 and fell to zero on test day 3. The lower panels present the data from all of the test trials on each of the three days of training, and indicate that the greater discrimination between AX and BX in later test sessions was accompanied by decreased potentiation to B, the presumed conditioned inhibitor, as well as decreased potentiation to AB.

Statistical analyses supported these observations. In the omnibus ANOVA, only the main effect of test cue reached signifi-
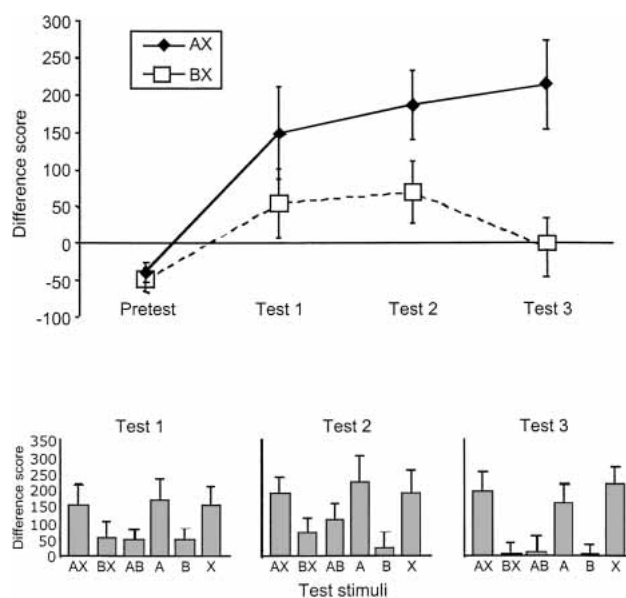

Figure 7 *Top:* Mean startle difference scores obtained on AX and BX test trial types in the pretest and posttests of Experiment 4. Tests 1, 2, and 3 involved separate groups of animals. *Bottom* panels: Mean startle difference scores obtained on all test trial types in the posttests of Experiment 4.

cance [F(1,35) = 15.49; P < 0.01]. However, separate repeated-measures ANOVAs comparing potentiation to AX and BX on each test day revealed that the difference was significant on test 3 [F(1,35) = 11.44; P < 0.01] but not on test 1 or test 2 [F's < 3.51; P's > 0.05]. This pattern is entirely consistent with the predicted course of the discrimination presented in Figure 4, which involves responding to both AX and BX in the early stages of the discrimination and little to no responding to BX in later stages.

Interestingly, the Rescorla-Wagner model (1972) predicts an up-and-down pattern of responding to BX because the increase in excitation to X occurs somewhat more rapidly than does the accrual of inhibition to B. Thus, responding to BX returns to zero when B becomes sufficiently inhibitory to counteract responding to X. Because the AB summation tests in Experiments 3 and 4 indicated strong inhibition of responding to A by B, it seems likely that the up-and-down pattern of responding to BX does indeed reflect the development of inhibition to B. Given that the difference between AX and BX was not statistically reliable until after the third training session, each of which involved five presentations each of AX+ and BX−, it would seem that 10–15 presentations of these cues are necessary for inhibition to develop completely.

DISCUSSION

The present series of experiments had two goals: first, to examine the conditions under which fear-potentiated startle is externally inhibited, and second, to develop a conditioned inhibition training protocol for use in this paradigm that minimizes the contribution of external inhibition to response suppression. Experiment 1 demonstrated that the magnitude of external inhibition of fear-potentiated startle is related systematically to the temporal placement of the external inhibitor with respect to the excitor in test, such that simultaneous compounds of these two stimuli produce maximal response suppression, whereas zero-trace and trace arrangements are associated with smaller decrements. This phenomenon is not due to an unconditioned suppression of baseline startle by the novel external inhibitor, but rather reflects a suppression of conditioned fear to the excitor.

Experiment 2 replicated the external inhibition effect with simultaneous compounds and demonstrated that this effect may be lessened in magnitude via pretest exposure to the excitor and external inhibitor in compound with other stimuli. Experience with the external inhibitor, either alone or in compound, prior to test was more effective than experience with the excitor as part of a compound, but both manipulations combined were maximally effective and actually eliminated external inhibition in a group that had the excitor trained in compound with a second stimulus and the to-be-inhibitor pre-exposed in compound with yet another stimulus (group AX+/X−/BY−). These findings are consistent with an attentional account of external inhibition, which proposes that the novel external inhibitor lessens the magnitude of the conditioned response by attracting attention away from the familiar excitor (Pavlov 1927). By this account, pre-exposure to the external inhibitor, combined with pre-exposure to stimulus compounds, would be expected to reduce the novelty of these cues and the magnitude of the orienting response to them. The fact that preexposure was effective in this experiment but not in prior work by Falls (1993) may have resulted from the use of a pure tone in this case as opposed to a white noise in his experiments, although a head-to-head comparison of these two types of auditory stimuli would be necessary to test this directly.

Regardless of the mechanism involved, the fact that external inhibition of fear-potentiated startle may be attenuated through certain procedural manipulations in training is fortuitous in the sense that incorporating such manipulations into a conditioned

inhibition design may lessen the problem of external inhibition in summation tests. Experiments 3 and 4 validated one such procedure, the AX+, BX− discrimination, in which the contribution of external inhibition to fear inhibition was diminished via pre-exposure to the external inhibitor and to stimulus compounds. Experiment 3 demonstrated that the AX+, BX− discrimination can be learned by rats in this paradigm and that the pattern of responsiveness to the individual cues (A, B, and X) and stimulus compounds (AX, BX, and AB) is remarkably similar to that predicted by the Rescorla-Wagner model (1972; Wagner and Rescorla 1972). Critically, fear-potentiated startle to AB was less than that to A from the first presentations of these cues in test, indicating that B did indeed acquire conditioned inhibition. Further substantiating B's status as a conditioned inhibitor was the up-and-down pattern of responding to BX over the course of the discrimination (Experiment 4), which the Rescorla-Wagner model attributes to the gradual accrual of inhibition to B and subsequent counteracting of excitation to X.

In principle it might be argued that it would be simpler, and perhaps even better, to study fear inhibition through the use of a traditional A+, BA− procedure. Despite this, we believe that the AX+, BX− discrimination offers some significant advantages beyond its circumventing of external inhibition. In particular, the AX+, BX− discrimination seems to render the B stimulus to be mostly inhibitory, unlike the serial conditioned inhibition design developed by Falls (1993) and Falls and Davis (1997). In their experiments B did acquire inhibition, as evidenced by its ability to suppress responding to A and to a separately trained C stimulus when the onset of those target stimuli co-occurred with the offset of B, but B also developed significant excitation, as evidenced by its ability to potentiate startle elicited in its presence. There are several possible explanations for this outcome, of which two—occasion setting and second-order conditioning— seem most likely.

Occasion setting is an ability acquired by a CS to modulate (i.e., "set the occasion for;" Skinner 1938) responding to another CS by signaling the delivery or omission of reinforcement (for review, see Swartzentruber 1995). In general it is believed that serial (trace) compound stimulus presentations, including zero-trace, promote occasion setting to the first CS of the series (Ross and Holland 1981; Holland 1985), although occasion setting may develop under other circumstances as well (Holland 1989; Brandon and Wagner 1991). Among other features, occasion setters are distinguished from simple CSs by two properties: first, an occasion setter is capable of modulating responding to a target CS independently of its ability to elicit a CR on its own, as in the case of a stimulus that acts both as a simple excitor (eliciting a CR) and a negative occasion setter (inhibiting the CR that normally would be elicited by a target CS). Second, the modulatory abilities of occasion setters are resistant to disruption by direct reinforcement of negative occasion setters or nonreinforcement of positive occasion setters. The B stimulus in the Falls and Davis (1997) study exhibited both of these properties, in that it was able to potentiate startle elicited in its presence and inhibit startle to A when the onset of A coincided with the offset of B, and its inhibitory capacity was spared when B was reinforced at a delay interval. Thus, although Falls and Davis (1997) favored an alternative interpretation of their data, it is possible that the zero-trace compounds they employed promoted the acquisition of negative occasion setting abilities by B, which in turn allowed B to acquire simple excitatory tendencies through other mechanisms.

A major candidate mechanism for the acquisition of excitation by B in the Falls and Davis (1997) study is second-order conditioning, in which a neutral CS that is presented in compound with an excitatory CS itself becomes somewhat excitatory,

presumably because the excitatory (first-order) CS acts as an unconditioned stimulus (Rescorla 1980). Second-order conditioning is a puzzling phenomenon because it endows a CS with excitation and yet develops under the same training conditions as those most commonly used to promote the development of conditioned inhibition, that is, A+, BA−. Theories of second-order conditioning and its relation to conditioned inhibition generally are lacking (e.g., Moore and Stickney 1980), but it has been proposed that second-order excitation accrues during the early BA− trials whereas conditioned inhibition predominates with continued training (Rescorla 1980). Thus, although second-order conditioning may be a transient and relatively weak phenomenon, it nevertheless may have contributed to the startle-potentiating properties of B in the experiments of Falls and Davis (1997). In fact, Falls (1993) found no evidence for a decrease in excitation to B with continued training; instead, the requisite event for inhibition appeared to be the offset of B, which predicted the absence of the US. This was especially evident in an experiment in which the acquisition of excitation to B was retarded when B was paired with the US in a trace arrangement, that is, when US delivery occurred after the offset of B, at a point in time at which the US previously had been omitted.

Several properties of the AX+, BX− discrimination circumvent these problems. First, because simultaneous rather than serial compounds are employed, the potential for A and B to develop occasion-setting properties is minimized (Holland 1985). Second, because X never becomes fully excitatory (i.e., because it is partially reinforced and overshadowed by A on AX+ trials), it should be relatively ineffective as a first-order stimulus and should not endow B with measurable excitation in the BX− trials (i.e., it should be relatively ineffective in supporting second-order conditioning to B). These features, together with the elimination of external inhibition in test, make the AX+, BX− discrimination an attractive protocol for the study of the neural bases of conditioned fear inhibition. Particularly for molecular and genetic analyses, which often rely on technically difficult and labor-intensive techniques to detect differences in the expression of a small number of genes, it is critical to isolate as much as possible inhibitory learning processes from excitatory and modulatory tendencies whose underpinnings may be quite distinct.

The apparent ability of AX+, BX− to render B a conditioned inhibitor does come at a cost, however: relative to the more traditional A+, BA− conditioned inhibition training protocol, AX+, BX− discrimination training leaves B less strongly inhibitory than it otherwise might be. Thus, returning to the simulated outcome of the AX+, BX− discrimination based on the Rescorla-Wagner model (1972) presented in Figure 4, it is apparent that B attains a value of −0.33, meaning that it is about a third as inhibitory as it would have been if trained as A+, BA−. This aspect of the AX+, BX− procedure is its biggest liability, as one would certainly prefer that the magnitude of inhibition be as great as possible; however, in light of the arguments presented above in favor of a purely inhibitory B stimulus, we feel this is a secondary concern.

It could be argued that the results of Experiment 2 indicate that A+, BA− training could be salvaged with pre-exposure to the various cues. That is, if A+, BA− training were preceded by a pre-exposure phase involving, for example, presentations of AX+ and BY−, then one could proceed with the use of simultaneous compound stimulus presentations in the discrimination training and test phases and expect external inhibition on BA test trials to be minimized. Although this possibility is attractive in the sense that one would be left with a B stimulus that is maximally inhibitory (i.e., −1.0 as opposed to −0.33), it does nothing to address the problem of second-order conditioning to B. Moreover, because stimulus pre-exposure is in a sense "built in" to the

AX+, BX− procedure, it is not necessary to include a pretraining phase prior to acquisition of the discrimination. Finally, nonreinforced pretraining of B might introduce another problem, namely, latent inhibition to B, which would then retard B's acquisition of inhibitory properties (Rescorla 1971).

Accepting that the AX+, BX− discrimination is a valid and useful behavioral protocol, one may carry out cellular manipulations in various brain areas before training to examine their effect on the development of inhibition, before test to examine their effect on the expression of inhibition, or before both to address questions of state dependency. Among the issues to be considered in future experiments is the effect of medial prefrontal cortex (mPFC) inactivation on conditioned excitation and inhibition in the AX+, BX− paradigm, as several studies suggest an involvement of this structure in conditioned fear extinction but others have failed to replicate those findings (cf. Myers and Davis 2002). It will also be important to assess whether drugs that facilitate (e.g., D-cycloserine) or retard (e.g., AP5) extinction have similar effects on the development of inhibition (Falls et al. 1992; Walker et al. 2002). The AX+, BX− discrimination may also be useful in studies examining the molecular underpinnings of the development and/or expression of excitation and inhibition. For example, if one were to train animals on the discrimination and, sometime thereafter, expose separate groups to A, B, and AB, it should be possible, through the use of techniques such as gene arrays, in situ hybridization, and immunocytochemistry to compare patterns of gene expression and protein synthesis among the groups and tease apart any differences resulting from exposure to an excitatory, inhibitory, or mixed cue. Although these types of analyses have been applied to extinguished CSs (Chhatwalet al. 2003), the inherent confounding of excitation and inhibition within the extinction procedure makes it difficult, if not impossible, to examine changes resulting from inhibitory learning per se.

Finally, it is possible to examine the development and expression of the AX+, BX− discrimination in primates, and thereby probe the involvement of higher-order cortical structures, such as the mPFC, in species closely related to humans. We have examined the AX+, BX− discrimination in fear-potentiated startle with humans and obtained results remarkably similar to those described in this paper, including inhibition of potentiated startle on AB test trials relative to A test trials (Fiallos et al. 2003). It should now be possible to combine this paradigm with imaging technologies to identify brain areas activated by conditioned excitors, areas activated by conditioned inhibitors, and patterns of activation resulting from compound excitor-inhibitor presentations.

ACKNOWLEDGMENTS

REFERENCES

Bouton, M.E. 1993. Context, time and memory retrieval in the interference paradigms of Pavlovian conditioning. *Psychol. Bull.* **114:** 80–99.

———. 2000. A learning theory perspective on lapse, relapse, and the maintenance of behavior change. *Health Psychol.* (Suppl.) **19:** 57–63.

Brandon, S.E. and Wagner, A.R. 1991. Modulation of a discrete Pavlovian conditioned reflex by a putative emotive Pavlovian conditioned stimulus. *J. Exp. Psychol. Anim. Behav. Process* **17:** 299–311.

Brown, J.S., Kalish, H.I., and Farber, I.E. 1951. Conditional fear as revealed by magnitude of startle response to an auditory stimulus. *J. Exp. Psychol.* **41:** 317–328.

Campeau, S. and Davis, M. 1992. Fear potentiation of the acoustic startle reflex using noises of various spectral frequencies as conditioned stimuli. *Anim. Learn. Behav.* **20:** 177–186.

Cassella, J.V. and Davis, M. 1986. The design and calibration of a startle measurement system. *Physiol. Behav.* **36:** 377–383.

Chhatwal, J.P., Myers, K.M., Ressler, K.J., and Davis, M. 2003. Dynamic regulation of gephyrin within the basolateral amygdala following extinction of conditioned fear. Program No. 426.9 2003 Abstract Viewer/Itinerary Planner. Washington, D.C., Society for Neuroscience.

Davis, M. 2000. The role of the amygdala in conditioned and unconditioned fear and anxiety. In *The Amygdala* (ed. J.P. Aggleton), Vol. 2, pp. 213–287. Oxford University Press, Oxford, UK.

Davis, M. and Astrachan, D.I. 1978. Conditioned fear and startle magnitude: Effects of different footshock or backshock intensities used in training. *J. Exp. Psychol. Anim. Behav. Process* **4:** 95–103.

Davis, M., Schlesinger, L.S., and Sorenson, C.A. 1989. Temporal specificity of fear-conditioning: Effects of different conditioned stimulus-unconditioned stimulus intervals on the fear-potentiated startle effect. *J. Exp. Psychol. Anim. Behav. Process* **15:** 295–310.

Falls, W.A. 1993. "Conditioned inhibition of fear-potentiated startle." Ph.D. thesis, Yale University, New Haven, CT.

Falls, W.A. and Davis, M. 1994. Fear-potentiated startle using three conditioned stimulus modalities. *Anim. Learn. Behav.* **22:** 379–383.

———. 1997. Inhibition of fear-potentiated startle can be detected after the offset of a feature trained in a serial feature negative discrimination. *J. Exp. Psychol. Anim. Behav. Process* **23:** 3–14.

Falls, W.A., Miserendino, M.J.D., and Davis, M. 1992. Extinction of fear-potentiated startle: Blockade by infusion of an NMDA antagonist into the amygdala. *J. Neurosci.* **12:** 854–863.

Fiallos, A.M., Jovanovic, T., Keyes, M., Myers, K.M., Davis, M., and Duncan, E. 2003. Cognitive awareness versus potentiated startle evidence of learning in a fear discrimination paradigm in humans. Poster presented at the *10th Annual Meeting of the Cognitive Neuroscience Society*, New York.

Gewirtz, J.C., Falls, W.A., and Davis, M. 1997. Normal conditioned inhibition and extinction of freezing and fear potentiated startle following electrolytic lesions of medial prefrontal cortex. *Behav. Neurosci.* **111:** 712–726.

Holland, P.C. 1985. The nature of conditioned inhibition in serial and simultaneous feature negative discriminations. In *Information processing in animals: Conditioned inhibition* (eds. R.R. Miller and N.E. Spear), pp. 267–298. Lawrence Erlbaum, Hillsdale, NJ.

———. 1989. Occasion setting with simultaneous stimulus compounds in rats. *J. Exp. Psychol. Anim. Behav. Process* **15:** 183–193.

Konorski, J. 1967. *Integrative activity of the brain: An interdisciplinary approach*. The University of Chicago Press, Chicago.

LeDoux, J.E. 1996. *The emotional brain: The mysterious underpinnings of emotional life*. Simon and Schuster, New York.

Lovibond, P.F., Davis, N.R., and O'Flaherty, A.S. 2000. Protection from extinction in human fear conditioning. *Behav. Res. Ther.* **38:** 967–983.

Moore, J.W. and Stickney, K.J. 1980. Formation of attentional-associative networks in real time: Role of the hippocampus and implications for conditioning. *Physiol. Psychol.* **8:** 207–217.

Myers, K.M. and Davis, M. 2002. Behavioral and neural analysis of extinction. *Neuron* **36:** 567–584.

Paschall, G.Y. and Davis, M. 2002. Olfactory-mediated fear-potentiated startle. *Behav. Neurosci.* **116:** 4–12.

Pavlov, I.P. 1927. Conditioned reflexes (G.V. Anrep, Trans.). Oxford University Press, London.

Reiss, S. and Wagner, A.R. 1972. CS habituation produces a "latent inhibition effect" but no active "conditioned inhibition." *Learn. Motiv.* **3:** 237–245.

Rescorla, R.A. 1969. Pavlovian conditioned inhibition. *Psychol. Bull.* **72:** 77–94.

———. 1971. Summation and retardation tests of latent inhibition. *J. Comp. Physiol. Psychol.* **75:** 77–81.

———. 1980. *Pavlovian second-order conditioning: Studies in associative learning*. Lawrence Erlbaum, Hillsdale, NJ.

———. 2001. Experimental extinction. In *Handbook of contemporary learning theories* (eds. R.R. Mowrer and S.B. Klein), pp. 119–154. Lawrence Erlbaum, Mahway, NJ.

Rescorla, R.A. and Wagner, A.R. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning II: Current research and theory* (eds. A. Black and W.F. Prokasy), pp. 64–99. Appleton-Century-Crofts, New York.

Richardson, R., Hess, M., and Campbell, B.A. 1994. The orienting response to brief auditory stimuli in preweanling and adult rats. *Dev. Psychobiol.* **27:** 93–100.

Rickert, E.J., Lorden, J.F., Dawson, R.D., Smyly, E., and Callahan, M.F. 1979. Stimulus processing and stimulus selection in rats with hippocampal lesions. *Behav. Neural. Biol.* **27:** 454–465.

Ross, R.T. and Holland, P.C. 1981. Conditioning of simultaneous and serial feature-positive discrimination. *Anim. Learn. Behav.* **9:** 293–303.

Skinner, B.F. 1938. *The behavior of organisms*. Appleton-Century-Crofts, New York.

Swartzentruber, D. 1995. Modulatory mechanisms in Pavlovian conditioning. *Anim. Learn. Behav.* **23:** 123–143.

Vouimba, R.M., Garcia, R., Baudry, M., and Thompson, R.F. 2000. Potentiation of conditioned freezing following dorsomedial prefrontal cortex lesions does not interfere with fear reduction in mice. *Behav. Neurosci.* **114:** 720–724.

Wagner, A.R. and Rescorla, R.A. 1972. Inhibition and learning. In *Inhibition in Pavlovian conditioning* (eds. R.A. Boakes and M.S. Halliday), pp. 301–335. Academic Press, London.

Wagner, A.R., Logan, F.A., Haberlandt, K., and Price, T. 1968. Stimulus selection in animal discrimination learning. *J. Exp. Psychol.* **76:** 171–180.

Walker, D.L., Ressler, K.J., Lu, K.-T., and Davis, M. 2002. Facilitation of conditioned fear extinction by systemic administration or intra-amygdala infusions of D-cycloserine as assessed with fear-potentiated startle in rats. *J. Neurosci.* **22:** 2343–2351.