Overshadowing and Associability Change

Peter M. Jones and Mark Haselgrove
University of Nottingham

Three appetitive Pavlovian conditioning experiments with rats examined the associability of stimuli A and B that had a history of compound conditioning (AB+), relative to stimuli X and Y that had a history of conditioning in isolation (X+, Y+). Following this training, Experiment 1 revealed that conditioned responding was higher to X and Y than to A and B (overshadowing). In a subsequent AY+, AX−, BY− test discrimination, the AY/BY discrimination was solved more readily than the AY/AX discrimination. In Experiment 2, following AB+, X+, Y+ training, A and Y were presented as a compound and signaled the availability of reinforcement upon the performance of an instrumental response. Test trials in which A and Y were presented alone, and in extinction, revealed that A acquired greater control of instrumental responding than Y. Experiment 3 revealed that following AB+, X+, Y+ training, A and B served as more effective discriminative stimuli for instrumental responding than X and Y. Overall, these results imply that the associability of stimuli conditioned in compound is higher than stimuli conditioned in isolation. These results are discussed in terms of attentional theories of associative learning.

Keywords: overshadowing, cue competition, associability, attention

In a typical overshadowing experiment, two conditioned stimuli (CSs) are presented together as a simultaneous compound and followed, for a number of trials, by an unconditioned stimulus (US). In a subsequent test, the conditioned response to one of these CSs is revealed to be weaker than to the compound of the CSs, or to a control CS that has been presented in isolation and paired with the US. Pavlov (1927, p.141) was the first to provide an experimental demonstration of overshadowing. In his experiment with dogs, a thermal stimulus of 0 °C and tactile stimulation of the skin were presented simultaneously and followed by consumption of a dilute acid. The conditioned response of salivation was substantial during the presentation of the compound, yet when the thermal CS was presented alone, no salivation was detected—learning about the thermal CS was said to have been overshadowed by the tactile CS. Overshadowing is a robust property of Pavlovian conditioning, and has been demonstrated in rats in studies of conditioned suppression (Kamin, 1968; Mackintosh, 1971), flavor-aversion and flavor-preference learning (Dwyer, Haselgrove, & Jones, 2011; Revusky, 1971), appetitive conditioning (Holland, 1999) and spatial learning (Pearce, Graham, Good, Jones, & McGregor, 2006). It has also been demonstrated in a diverse range of species from goldfish (Tennant & Bitterman, 1975) to humans (Chamizo, Aznar-Casanova, & Artigas, 2003).

Two classes of explanation for overshadowing can be derived from elemental theories of associative learning, which differ in the emphasis that they place on the processing of the US or the processing of the CS on each conditioning trial. According to US-processing theories of learning (e.g., Rescorla & Wagner, 1972) the US can support a limited amount of associative strength (λ) and this limit is shared between the CSs that precede the US on each trial. Consequently, conditioning trials with a compound of two CSs will result in each CS acquiring less associative strength (V) than if they had been conditioned in isolation. According to CS-processing theories of learning, overshadowing is a consequence of the CS that has been conditioned in compound having a lower associability than a CS that has been conditioned in isolation (e.g., Mackintosh, 1975; Sutherland & Mackintosh, 1971). According to Mackintosh's (1975) theory, for example, the associability (α) of a CS increases if it is a better predictor of the US than all other CSs present on a trial, and decreases if it is no better a predictor of the US than all the other CSs present on a trial. The rules specified by Mackintosh for determining these increases and decreases to a target CS (A) are shown in Equations 1a and 1b:

$$\Delta\alpha_A > 0 \text{ if } |\lambda - V_A| < |\lambda - V_X| \tag{1a}$$

$$\Delta\alpha_A < 0 \text{ if } |\lambda - V_A| \geq |\lambda - V_X| \tag{1b}$$

Where V_X is the sum of the associative strength of all available CSs, $-V_A$. The size of the change in α is assumed to be proportional to the magnitude of the inequalities in Equation 1a and 1b. When a compound of two stimuli is paired with a US the associability of the target (overshadowed) CS will either remain unchanged (if the target and the overshadowing CS initially have equal associabilities) or fall (if the target CS possesses a lower initial associability than the overshadowing CS); either of these circumstances may limit the extent to which the target CS can enter into associations with the US. In contrast, a CS that is presented in isolation and paired with a US is the best predictor of the US; therefore, attention to this CS will increase, enabling it to enter into an association with the US.

CS- and US-processing theories of learning therefore, provide rather different accounts of the mechanisms responsible for overshadowing. Despite this, there has been relatively little recent study of which analysis provides the better account of this phenomenon. In a series of papers in the 1970s, Mackintosh (1971, 1976; Mackintosh & Reese, 1979) examined the properties of overshadowing and provided some evidence that it has, as its basis, a CS-processing mechanism. Mackintosh (1976) showed that two CSs would overshadow each other if, initially, they both had equivalent and low associability. If, initially, the two CSs both had equivalent and high associability, however, no overshadowing was obtained. These results are incompatible with the Rescorla-Wagner model (Rescorla & Wagner, 1972, pp. 79–80), which anticipates that conditioning with an AB compound will result in the associative strength of A being equivalent to the ratio $\alpha_A/(\alpha_A + \alpha_B)$, which implies that overshadowing should be observed whether the CSs used are both of a high or both of a low associability. This result can be understood from the perspective of Mackintosh's (1975) theory, however. Note that changes in α are assumed to be proportional to the difference between $|\lambda - V_A|$ and $|\lambda - V_X|$. During compound conditioning with CSs that enter the experiment with equivalent, low associability, therefore, the associability of these two CSs will remain unchanged at this low value, limiting conditioning to either CS. However, the associability of a CS that begins with low associability, and which is conditioned in isolation, can rise (as it is the best predictor of the US), thus permitting conditioning to this CS, and hence, overall, generating overshadowing. In contrast, if compound conditioning is conducted with CSs that begin the experiment with equivalent high associability then, according to the Mackintosh model, the associability of these two CSs will remain unchanged at this high value, permitting conditioning to both CSs and attenuating any overshadowing effect. As Mackintosh (1976, p. 192) notes "In the limiting case, where $\alpha_A = \alpha_B = 1$, there would be no reciprocal overshadowing at all."

Mackintosh's studies were not all so straightforwardly compatible with a CS-processing theory of learning, however. Mackintosh (1971; see also, James & Wagner, 1980; Mackintosh & Reese, 1979; Revusky, 1971) showed that overshadowing can be observed after only one trial of conditioning.[1] This result is difficult to reconcile with a CS-processing theory of learning such as that proposed by Mackintosh (1975), for it follows from this theory that at least two conditioning trials are required before the animal learns that neither CS is a better predictor of the US, and hence learns to ignore these CSs. This result is also difficult to reconcile with the Rescorla-Wagner model. According to this theory, after only one conditioning trial, the amount of associative strength acquired by a CS conditioned in compound with another CS should be equal to the associative strength acquired by a CS conditioned in isolation.

To the best of our knowledge there remains no evaluation of the central prediction of Mackintosh's (1975) account of overshadowing, which is that the associability of a CS conditioned as part of a compound should be, at best, equivalent to a CS conditioned in isolation[2]; and at worst, be lower than a CS conditioned in isolation. This omission presumably stems from the difficulty in assessing the associability of CSs that have very different levels of associative strength. However, in a recent series of conditioning experiments conducted with pigeons and rats, we have circumvented this problem and have demonstrated a number of techniques that permit the evaluation of the associability of a CS (Haselgrove, Esber, Pearce, & Jones, 2010; Pearce, Esber, George, & Haselgrove, 2008). For example, Haselgrove et al. gave rats two feature-positive discriminations of the form: AX+, BY+, X−, Y−. According to Mackintosh's (1975) theory, on the AX+ and BY+ trials, A and B are better predictors of the US than are X and Y, which themselves are only paired with the US on 50% of the trials. This training should therefore result in A and B gaining more associability than X and Y. To test this prediction, Haselgrove et al. subsequently gave the rats an AY+, AX−, BY− test discrimination. If the associability of A and B is higher than the associability of X and Y then the subdiscrimination between AY and BY (which is based upon two stimuli that should possess high associability) should be solved more readily than the subdiscrimination between AY and AX (which is based upon two stimuli that should possess low associability). This is exactly what Haselgrove et al. found. Crucially, each compound in the test discrimination is composed of one CS that has high associative strength and one CS that has low associative strength. According to US-processing theories of learning such as the Rescorla-Wagner model, therefore, the AY/AX and AY/BY subdiscriminations should progress at the same rate. The goal of the current experiments was to use this, and other techniques to assess the associability of CSs that have a history of compound conditioning (AB+) or of conditioning in isolation (X+, Y+). Our expectation was that the results would confirm the predictions of Mackintosh's theory and show that the associability of A and B was lower than the associability of X and Y.

Experiment 1

During stage 1, rats received appetitive Pavlovian conditioning in which a compound of two stimuli was followed on each trial by the delivery of food (AB+); in addition, trials were included in which two stimuli were presented in isolation and also paired with the delivery of food (X+, Y+). To discourage the rats from learning simply that presentations of all stimuli are followed by food, nonreinforced trials with an additional stimulus were also included (P−). Probe trials conducted in extinction for animals in Group Test examined the strength of responding to A, B, X, and Y. We anticipated a standard overshadowing effect, that is to say, stronger responding to X and Y than to A and B. If overshadowing is demonstrated, then Mackintosh's theory is constrained to suggesting that the associability of A and B is lower than the associability of X and Y. This prediction was assessed in a subsequent test in which all rats were given an AY+, AX−, BY− discrimination. Under these circumstances, Mackintosh's theory predicts that the subdiscrimination between AY and AX should progress more readily than the subdiscrimination between AY and BY. It is possible that the overshadowing test, conducted in extinction, with

---

[1] In fact, both Mackintosh (1971) and Mackintosh and Reese (1979) showed one-trial overshadowing occurred when the test (overshadowed) CS was considerably less salient than the overshadowing stimulus. When these two stimuli were better matched in terms of salience, one-trial overshadowing was not observed. This distinction was not observed by James and Wagner (1980).

[2] Of course, under these circumstances, Mackintosh's (1975) theory predicts no overshadowing either.

A, B, X, and Y in Group Test may have differentially influenced the associability of A and B relative to X and Y before the discrimination stage. To address this possibility, Experiment 1 also included an additional group (Group No-Test) that was treated in an identical fashion to Group Test, with the exception that the overshadowing test trials with A, B, X, and Y were omitted.

Method

Subjects. Thirty-two naïve, male Lister hooded rats (*Rattus norvegicus*) supplied by Harlan Olac (Bicester, Oxon, England) served as experimental subjects. They were housed in pairs in a light-proof holding room in which the lights were on for 12 hr/day. At the time of their arrival their weight range was 250–300 g. Before the start of the experiment, they were gradually reduced to 80% of their free-feeding weights; during the experiment, rats' weights were allowed to increase in proportion with a separate group of rats housed in the same room who were allowed free access to food, to allow for healthy growth. They were maintained at these weights throughout the experiment by being fed a restricted amount of food after each experimental session. Half of the rats were randomly assigned to Group Test, while the remaining rats were assigned to Group No-Test. Experimental sessions were conducted at the same time each day, during a period when the lights were on in their holding room.

Apparatus. Eight conditioning chambers (MED Associates, St. Albans, VT; 30.0 × 24.0 × 20.5 cm high) were housed in separate light- and sound-attenuating boxes. Exhaust fans in each box provided ventilation and a background noise of 70 dB. The two smaller walls of each chamber were constructed from aluminum, the two larger walls (one of which served as a door) and the ceiling were made of clear acrylic, and the floor was composed of a series of 19 stainless steel rods of 4.8-mm thickness, positioned 1.6 cm apart and parallel to the shorter walls. One of the shorter walls was equipped with a square-shaped recessed magazine, 50 mm wide, to which 45-mg food pellets (traditional formula, P. J. Noyes, Lancaster, NH) could be delivered. The magazine was located 18 mm above the grid floor and equidistant from the two adjacent walls. An infrared beam was sent from one lateral side of the magazine and received on another. Interruption of this beam could be detected and recorded as magazine activity. Each chamber was also equipped with two retractable levers (MED Associates) located on either side of the magazine. These levers could be either presented to the rat or fully retracted into the wall of the chamber, a state that was controlled by the experimental computer. When presented, the levers protruded 1.9 cm from the wall, and were 4.8 cm wide and 0.2 cm thick. The horizontal distance from the midpoint of the wall to the midpoint of each lever was 7.9 cm, and each lever was 6.2 cm from the grid floor. For this experiment the levers remained retracted throughout each experimental session. The opposite wall of the chamber incorporated three loudspeakers: two square-shaped loudspeakers of 70-mm width located in the upper left and upper right corners of the wall, and a 35 × 70 mm loudspeaker located 35 mm from the ceiling and equidistant from the two adjacent walls. The upper left loudspeaker could emit a pulsed 77-dB, 8-kHz tone, with each 0.25-s pulse separated by a 0.25-s silent period. This speaker could also emit a 74-dB "buzzer" composed of a 500-Hz train of clicks. The upper right loudspeaker could emit a white noise with an amplitude of 78 dB. The rectan-

gular loudspeaker could emit a 77-dB, 2.9-kHz tone. A relay was attached to the outside of this wall, and could produce a 76-dB train of clicks such that a click occurred every 0.08 s. Five auditory stimuli could therefore be produced (a pulsed 8-kHz tone, a 2.9-kHz tone, a white noise, a clicker and a buzzer), each of which was 15 s in duration. Where two stimuli were presented in compound, they had simultaneous onset and offset. Control of experimental events and recording of magazine activity was conducted by an attached computer that was programmed with the MED-PC programming language.

Procedure. Before the beginning of the experiment, auditory stimuli were assigned to serve as stimuli A, B, X, and Y. For half the rats within each group, the white noise and 2.9-kHz tone were stimuli A and B; for the remaining half, the clicker and the pulsed 8-kHz tone were A and B. Within each of these two subgroups of rats, stimuli assigned to A and B were counterbalanced, as were stimuli assigned to X and Y. For all rats, the buzzer served as stimulus P. All rats initially received a single session of magazine training and during this 30-min session, one pellet was delivered to the magazine every 60 s. All rats successfully consumed all pellets during this session.

During each of the following 12 sessions, all rats received AB+, X+, Y+, P− training. Each session was 52 min long, with four trials of each type distributed in a block-randomized manner such that two trials of each type were contained in each successive block of eight trials. The intertrial interval (ITI), defined as the period from the end of one trial to the start of the next, varied randomly from 65–145 s, with a mean ITI for each session of 105 s. Presentations of the compound stimulus, AB, and the individual stimuli X and Y, were immediately followed by the delivery of two food pellets to the magazine. Following these six sessions, a test session was carried out to assess the extent of overshadowing. This session was composed of four of each kind of training trial as described above, in addition, animals in Group Test received two nonreinforced test trials with each of A, B, X, and Y. These trials were arranged into six blocks of four trials, with trials within each block occurring in a random sequence; the first, second, fifth and sixth blocks contained one of each type of training trial, and the third and fourth blocks contained one of each type of nonreinforced test trial. Rats in Group No-Test received a session that was identical to this with the exception of the test trials, which were omitted; the structure of the test session, including the temporal position of the training trials, was otherwise the same for the two groups.

During the final five sessions of the experiment, rats received training with an AY+, AX−, BY− discrimination. Each session in this stage was 68 min in duration, and contained 16 trials with AY+ and eight of each of the other two trial types. Each successive block of eight trials contained four trials with AY+ and two each with AX− and BY−, in a random order. Other details of these sessions were the same as for the earlier training sessions.

Results and Discussion

A type I error rate of $p < .05$ was adopted for all of the statistical tests in this and the subsequent experiments. Conditioning in stage 1 proceeded smoothly for both Group Test and No-Test, and by the final session of conditioning, the mean number of responses per minute made by Group Test during P, the AB compound and

during the average of X and Y were, respectively, 6.6, 40.5, and 36.1. The same means for Group No-Test were 8.3, 47.2, and 47.7. A two-way analysis of variance (ANOVA) with the variables of group (Test vs. No-test) and CS (AB, X/Y, and P) revealed an effect of CS, $F(2, 60) = 75.11$, $MSE = 178.68$, but no effect of group, $F(1, 30) = 2.22$, $MSE = 479.31$, and no interaction between these variables, $F(2, 60) = 1.08$, $MSE = 178.68$. The data were therefore collapsed across the group variable and post hoc tests conducted according to the Bonferroni procedure were performed to examine where the differences among the CSs lay. These comparisons revealed a difference between AB and P, and between XY and P, $ts(31) > 8.94$, but no difference between AB and XY, $t(31) = 1.01$. The number of responses made during the 15-s period before the onset of the CSs for Group Test and Group No-Test were, 4.7 and 5.3, respectively. The difference between these means was not significant, $t(30) = 0.39$.

During the test for overshadowing the mean number of responses per minute made by Group Test during the average of A and B and during the average of X and Y were 19.5 and 28.4, respectively. The difference between these means was significant, $t(15) = 4.53$. The mean number of responses made during the 15-s periods before the test trials during the overshadowing test was 3.2.

The top left- and top right-hand panels of Figure 1 show the mean number of responses per minute made during AY, AX, and BY across the five sessions of discrimination testing for Group Test and Group No-Test. Also shown on this figure is the number of responses per minute made during the 15-s intervals before these trials for the two groups. Overall, performance is slightly greater in Group No-Test than in Group Test, presumably reflecting the effects of the extinction trials with A, B, X, and Y in the latter group. However, surprisingly, the AY/BY discrimination was solved more readily than the AY/AX discrimination. This effect was also demonstrated in the group that did not receive a test for overshadowing before the test discrimination. To simplify the analysis of these data, discrimination ratios were calculated:
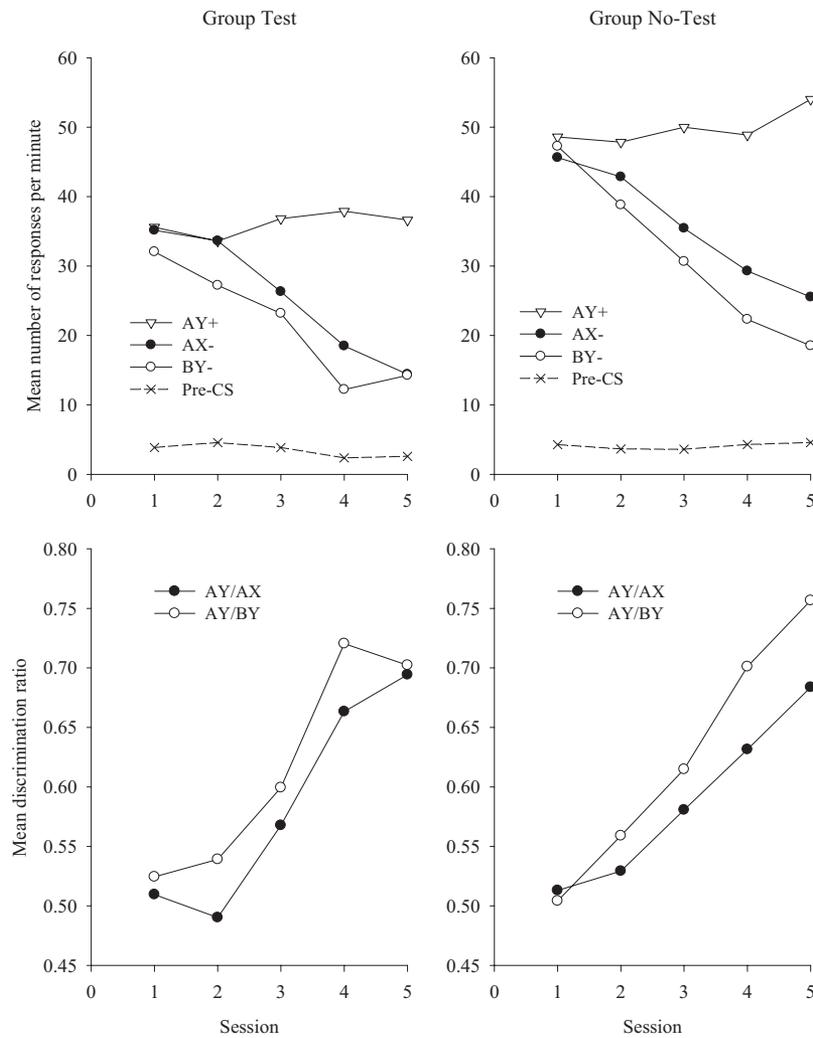

Figure 1. Results of the test discrimination from Experiment 1. The top left- and right-hand panels show the mean number of responses per minute during AY, AX, BY, and during the pre-CS intervals across the five test sessions for Group Test and Group No-Test respectively. + and − refer to food and no-food, respectively. The bottom panel shows discrimination ratios calculated from these means for the same groups.

p/(p + q) in which p is the number of responses during AY and q is the number of responses during AX or BY. A discrimination ratio of .5 indicates no discrimination between the compounds, and a discrimination ratio of 1 indicates a perfect discrimination. These discrimination ratios for the AY/AX and AY/BY discriminations are shown for Group Test and Group No-Test in the bottom left- and right-hand panel of Figure 1. A three-way ANOVA of individual discrimination ratios with the variables of group (Test vs. No-Test), discrimination (AY/AX vs. AY/BY), and session (1 to 5) was conducted. This analysis revealed no effect of group, $F < 1$, $MSE = 0.047$, but effects of session, $F(4, 120) = 62.27$, $MSE = 0.008$, and importantly, of discrimination, $F(1, 30) = 4.23$, $MSE = 0.020$. The three-way interaction and none of the two-way interactions were significant, $F(1, 30) < 1$ and $Fs(4, 120) < 1.74$, $MSEs > 0.004$ and < 0.020. Although not strictly permitted in the absence of a significant three-way interaction, it is nonetheless important to confirm that the two discriminations were solved at different rates in Group No-Test. Simple effects analysis confirms this to be the case, as the discrimination between AY and BY was greater than the discrimination between AY and AX for Group No-Test on Sessions 4 and 5, $Fs(1, 150) > 4.39$, $MSE = 0.007$. The mean numbers of responses per minute made during the pre-CS periods were averaged across the five sessions of the test discrimination. The mean number of responses per minute made during this period in Group Test and Group No-Test were 3.4 and 4.1, respectively. The difference between these means was not significant, $t(30) = 0.62$.

Following conditioning in which a compound of A and B was paired with food and trials in which X and Y were individually paired with food, rats made more responses during X and Y than during A and B, an overshadowing effect. According to Mackintosh's (1975) theory, overshadowing is a consequence of a loss of associability to A and B and/or a gain in associability to X and Y. Consequently, this theory predicts that the subsequent subdiscrimination between compounds that differ by the overshadowed CSs (AY and BY) should have been more difficult to solve than the subdiscrimination between compounds that differ by the CSs conditioned in isolation (AY and AX). This prediction was not confirmed. In fact, the opposite result was obtained: rats solved the discrimination between AY and BY more readily than the discrimination between AY and AX.

The results of Experiment 1 were not anticipated. Our expectation was in line with the prediction of Mackintosh's (1975) theory, that is to say, the associability of the overshadowed CSs would in fact be less than the associability of the CSs conditioned in isolation. Given the striking disparity between prediction and outcome, and the success that Mackintosh's theory has had elsewhere (for a review see Le Pelley, 2004) it is worth considering alternative reasons for why the results of this experiment were obtained. It is conceivable that AB+, X+, Y+ training did in fact result in the associability of A and B being less than the associability of X and Y, but this effect was masked, and indeed reversed by an additional process.

One candidate for this alternative process is based on a hypothesis proposed by Hall (2003). To explain the results of a number of perceptual learning studies (e.g., Blair & Hall, 2003a, 2003b), Hall proposed an antihabituation mechanism which opposes the normal habituation mechanism that is assumed to operate whenever any stimulus is frequently presented. According to this idea,

the perceptual effectiveness of a stimulus declines whenever it is presented (habituation). However, an antagonistic mechanism operates to restore the perceptual effectiveness of a stimulus if its representation is activated in the absence of that stimulus itself (antihabituation). Although the precise conditions that favor the operation of this mechanism have yet to be determined (Dwyer & Honey, 2007; Hall & Rodriguez, 2009), it nonetheless provides an interesting analysis of the effects of stimulus exposure on learning which can also be instantiated, in a slightly different form, in a more complex associative framework (McLaren & Mackintosh, 2000). During the tests for overshadowing in this experiment, A and B are presented in isolation. If within-compound associations developed between A and B during AB+ trials (Speers, Gillan, & Rescorla, 1980), then according to Hall's theory, presentation of A during the test for overshadowing would activate a representation of B in its absence, and vice versa, restoring the perceptual effectiveness of A and B. As X and Y were never presented as a compound during training, there are no grounds for anticipating that the perceptual effectiveness of these stimuli will be restored during the overshadowing test, as neither CS will have its representation evoked in its absence. If the restoration of the perceptual effectiveness of A and B was substantial, it is conceivable that it may have opposed the loss of associability to A and B that occurred during AB+ training. However, the solution of the subdiscrimination between the compounds that differed by a CS that had a history of compound conditioning (AY and BY) was superior to the solution of subdiscrimination between the compounds that differed by a CS that had been conditioned in isolation (AY and AX) even in Group No-Test, who did not receive any test trials with A, B, X, and Y. The results of this group therefore lend no support to the idea that the overshadowing test itself was responsible for the results of Experiment 1.

An additional possibility, however, is that an antihabituation process such as the one proposed by Hall (2003) is operative during the final AY+, AX−, BY− discrimination. If an association between A and B causes each of these stimuli to be associatively activated, and therefore perceptually enhanced during presentation of the other, then B will gain effectiveness during presentations of AY and AX, and A will gain effectiveness during presentations of BY, increasing the extent to which learning can take place with these stimuli. There will be no such enhancement for X and Y, and consequently the salience of A and B will be higher than that of X and Y. This advantage for A and B over X and Y will not be present from the outset of the final discrimination, but will develop as the discrimination proceeds and A and B have the opportunity to be associatively activated in their absence. The possibility remains, therefore, that AB+ training results in a loss of associability to A and B which is masked by a subsequent enhancement of their perceptual effectiveness that arises as a consequence of the discrimination testing procedure. This possibility is explored in Experiment 2.

Experiment 2

The main purpose of Experiment 2 was to address the possibility that the apparent increase in the associability of an overshadowed CS relative to a CS conditioned in isolation in Experiment 1 was because of the testing procedure used. According to theories that advocate the idea that learning results in variations in the process-

ing of the CS (e.g., Le Pelley, 2004; Mackintosh, 1975; Pearce & Hall, 1980; Wagner, 1978), associability is a property of the CS itself. As Mackintosh (1975, p. 277) notes, for example, ". . . the rate of conditioning to a particular stimulus is not just determined by some general learning rate parameter . . . but is also affected by a *stimulus-specific* parameter." In addition, contemporary analyses of stimulus associability frequently liken this concept with the extent to which a stimulus can attract attention. Beesley and Le Pelley (2011), for example, have shown that a stimulus that has undergone a loss of associability as a consequence of blocking training (Kamin, 1968) is also less able to attract overt attention, as measured by eye tracking (see also Kruschke, Kappenmann, & Hetrick, 2005; Wills, Lavric, Croft, & Hodgson, 2007). Consequently, variations in the associability of the CS should not only be revealed with a test that examines the subsequent rate of learning between CSs and a US, but also with different associative structures, such as those present in instrumental conditioning.

For the current experiment, rats again received Pavlovian conditioning in which an AB compound and X and Y were paired with food and P was not. After this training, an instrumental training stage was introduced in which rats received training such that responses on response 1 (R1), but not response 2 (R2) were reinforced during a compound of AY, and responses on R2 (but not R1) were reinforced during stimulus P. Finally, A and Y were presented individually over a series of test trials, and the rats were free to make responses on R1 and R2, in extinction. This training is summarized in Table 1. If the associability of, or attention to, A is greater than Y as a consequence of AB+, X+, Y+ training, then A should take greater control than Y over responding on R1 during the instrumental training stage. Thus at test, the difference between the rate of responding on R1 and the rate of responding on R2 should be greater for A than for Y. At no point in this experiment (the test of which is a variant on an optional-shift design; Duffaud, Killcross, & George, 2007) is B presented in the absence of A. Consequently, there is no opportunity for the perceptual effectiveness of A to be enhanced (Hall, 2003).

Method

Subjects and apparatus.   The subjects were 32 experimentally naïve male hooded Lister rats that were maintained in an identical fashion to the subjects in Experiment 1. The apparatus was the same as in Experiment 1.

Procedure.   The rats initially received one session of magazine training, the details of which were identical to Experiment 1. Rats were then trained to press the two levers during four sessions.

Table 1
Design of Experiment 2

Training	Instrumental training	Choice test
AB+		
X+	AY: R1+, R2−	A: R1 vs. R2
Y+	P: R1−, R2+	Y: R1 vs. R2
P−		

Note. A, B, X, Y, and P are all auditory cues; + and − refer to the presence and absence of food, respectively; R1 and R2 are two different instrumental responses. See text for further details.

Each of these sessions was 1 hr in length, and contained 40 1-min trials, separated by an ITI of 30 s. During each trial, a lever was presented and the rat allowed to earn single food pellets according to a reinforcement schedule that was reduced gradually from continuous reinforcement in the first session to a variable interval schedule with a mean interval of 15 s in the last. The left and right levers were each presented on half the trials, with the two trial types alternating. They were then trained for 12 sessions on an AB+, X+, Y+, P− discrimination in the same manner as for Experiment 1. Four instrumental conditioning sessions were then given with AY and P, with responding on R1 and R2 followed by the delivery of food during AY and P, respectively. Sessions in this stage were composed of 32 trials, each 20 s in length, with the same ITIs as used in training and a session length of 71 min. Each session contained 16 trials of each type, block-randomized such that each block of eight trials contained four of each type. During each trial, both levers were available; they were presented at the same time as the stimulus onset, and retracted at the termination of the stimulus. During the first 15 s of the trial, no lever responses resulted in the delivery of food to the magazine. Two pellets were delivered following the first "correct" response made during the last 5 s of the trial, with a correct response defined as a lever press made on R1 during AY, and R2 during P. For half the animals, the left lever served as R1 and the right lever served as R2, with the remaining animals receiving the opposite arrangement. This lever assignment was orthogonal to the assignment of stimuli. All instrumental responses made during the first 15 s of each trial were recorded by the computer. Because no food pellets were delivered during this time, this provides a measure of discrimination that is not compromised by the delivery of the US.

A test session was then given in which eight conditioning trials of each type were intermixed with eight test trials with each of A and Y. During each test trial, both levers were presented but no food was delivered. This test session failed to demonstrate control of responding by either A or Y, so four more conditioning sessions of the type already described were given.

Control of responding by A and Y was then assessed during two final test sessions. The first of these was 71 min in length, and contained an initial block of eight training trials, four each with AY and P in a random order, followed by 12 test trials with each of A and Y during which both levers were presented and no food delivered. These test trials were block-randomized such that each block of four trials contained two trials of each type. The second of these test sessions was 52 min in length, and contained 12 test trials of each type as described above but no training trials. Other details of these test sessions were the same as for the conditioning sessions already described.

Results and Discussion

By the final session of conditioning, one rat failed to provide any evidence of conditioning to AB (responses per minute <5), and throughout instrumental training, another rat made fewer than 50% correct responses during AY. These subjects were therefore omitted from all subsequent analyses.

Conditioning in stage 1, again, proceeded smoothly, and by the final session of conditioning, the mean number of responses per minute made during P, the AB compound, and during the average of X and Y were, respectively, 5.7, 35.3, and 32.8. A one-way

ANOVA revealed a difference among these means, $F(2, 58) = 96.03$, $MSE = 84.248$. Post hoc tests conducted according to Bonferroni procedure revealed the differences between AB and P, and X/Y and P, to be significant, $ts(29) > 9.80$, but the difference between AB and X/Y was not, $t(29) = 1.45$. The number of responses per minute made during the 15-s period before the onset of the CSs was 4.7. By the final session of instrumental conditioning the mean percentage of responses that were correct during the first 15 s of AY and P were, respectively, 76.0 and 74.6. Both of these means were significantly greater than chance (50%), $ts(29) > 8.96$, but they did not differ from each other, $t(29) = 0.31$.

During the first test session, the mean percentages of responses that were correct during the first 15 s of A (54%) and Y (56%) did not differ from chance (50%) for either stimulus, $ts(29) < 1.66$; furthermore, these means did not differ from each other, $t(29) = 0.36$. This result implies that four sessions of compound conditioning were not sufficient for either of these stimuli to take control over instrumental behavior. However, a further four conditioning sessions did result in the emergence of stimulus control of behavior—this control, however, was biased toward the CS that had a history of compound conditioning. The top panel of Figure 2 shows the mean percentage of responses that were correct during the first 15 s of A and Y across the three 4-trial blocks of the final two test sessions. By the end of both of these sessions, the mean percentage of correct responses was higher during A than Y, and furthermore there was a reduction in the mean percentage of correct responses during Y but not during A. A three-way ANOVA of individual percent correct responses with the variables of test (1 vs. 2), CS (A vs. Y) and trial block (1 to 3) confirmed these impressions. This analysis revealed no effects of test or CS, $Fs(1, 29) < 1.93$, $MSEs = 0.037$ and 0.154, but an effect of trial block, $F(2, 58) = 3.40$, $MSE = 0.023$. Neither the Test × CS nor the Test × Trial-block interactions, $Fs < 1$ $MSEs = 0.024$ were significant, and neither was the three-way interaction, $F(2, 58) = 1.51$, $MSE = 0.020$. Importantly, however, the CS × Trial-block interaction was significant, $F(2, 58) = 3.67$, $MSE = 0.020$. To illustrate the CS × Trial-block interaction, the bottom panel of Figure 2 shows the results collapsed across the two test sessions. Simple effects analysis of the CS × trial-block interaction revealed an effect of trial block for CS Y, $F(2, 116) = 6.76$, but not for A, $F < 1$, $MSEs = 0.021$. Furthermore there was a significant difference between A and Y on the final trial block, $F(1, 87) = 3.98$, but not on the earlier trial blocks, $Fs(1, 87) < 1.09$, $MSEs = 0.065$. These results imply two conclusions. First, the more appropriate responding during A than during Y suggests that the subjects learned more about A than Y during the preceding instrumental conditioning. Second, by the end of the test, Y had lost the ability to control the appropriate response, whereas A had not. This second conclusion is strengthened by the observation that on each of the (collapsed) trial-blocks, the mean percentage of correct responses was greater than chance during A, $ts(29) > 2.64$. However, the mean percentage of correct responses during Y was only greater than chance on the first trial block, $t(29) = 4.03$; after this point, responding was not significantly different from chance, $ts(29) < 1.32$.

Following instrumental lever press training, rats received trials in which an AB compound, X, and Y were paired with food and P was not. Subsequent instrumental training gave rats trials in which responses to R1 (but not R2) during an AY compound were
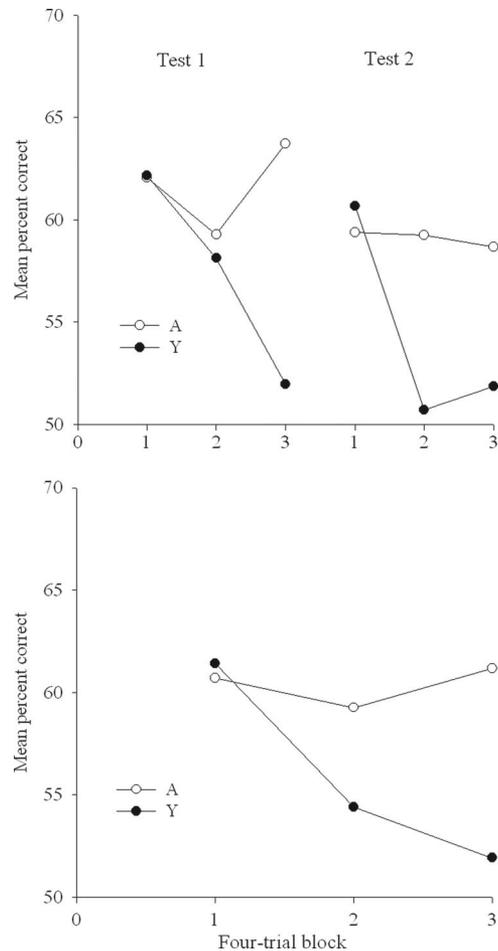

Figure 2. Results of the final choice tests from Experiment 2. The top panel shows, in four trial blocks, the mean percentage of correct responses that were made during A and Y during the two test sessions. The bottom panel shows the same data collapsed across the two tests.

reinforced, and responses to R2 (but not R1) during stimulus P were reinforced. Test trials in which a choice between R1 and R2 was provided during presentations of A and Y revealed that A had taken better control over responding than had Y. These results are consistent with the results of Experiment 1, and suggest that the associability of an overshadowed CS is greater than the associability of a CS conditioned in isolation: this poses something of a problem for the theory proposed by Mackintosh (1975), which makes the opposite prediction.

The results of this experiment are rather difficult to explain with the type of analysis that can be used to explain Experiment 1. First, there is no reason to suppose that in the current experiment, the representation of A was evoked (in its absence) more frequently than was the representation of Y. According to the theory proposed by Hall (2003), therefore, the perceptual effectiveness of these CSs should have been equivalent. An alternative possibility, however, is that the current experiment suggests that the associability of Y was, in fact, higher than A. In both choice tests, responding to A and Y was, at first, comparable; but with repeated nonreinforced test trials, stimulus Y lost the ability to control instrumental be-

havior, whereas A did not. That is to say, extinction was faster with Y than with A, which is consistent with the idea that the associability of Y was higher than the associability of A. There is nothing within Experiment 2 alone that allows us to refute this possibility, but consider the results of Experiment 1: if the associabilities of X and Y were higher than A and B, then, as we have already discussed, the results of this experiment would have been diametrically opposite to the results that we actually observed. However, the possibility remains that stage 1 training resulted in Y being better able to take control of instrumental responding than A. We address this issue in Experiment 3 by using a test procedure in which the potential for A and B to serve as discriminative stimuli for instrumental responding will be compared to that of X and Y.

Experiment 3

The purpose of Experiment 3 was to evaluate the claim that, following AB+ training, A and B have less potential to gain control of instrumental responding than stimuli conditioned in isolation. In stage 1, rats once again received conditioning trials in which an AB compound and X and Y were paired with food and P was not. Probe trials conducted in extinction with A, B, X, and Y established that conditioned responding to A and B was weaker than to X and Y. Finally, rats were trained on an instrumental discrimination in which performance of R1 during A or X was reinforced but R2 was not, and performance of R2 during B or Y was reinforced but R1 was not. This training is summarized in Table 2. If, as the results of Experiment 1 imply, the associability of, or attention to, A and B is more substantial than X and Y, then A and B should serve as better discriminative stimuli for instrumental responding than X and Y. Consequently, rats should learn to perform R1 during A, and R2 during B, with greater ease than during X and Y. Alternatively, if the results of the final test trials of Experiment 2 reflect faster extinction for Y than for A because attention to Y is higher, then we would expect the opposite pattern of results; the discrimination between A and B in the current experiment should be solved less readily than the discrimination between X and Y.

A further feature of this testing procedure is that a US can be obtained on each trial with A, B, X, and Y during the final instrumental discrimination. This has a benefit over the testing procedure used in Experiment 1, which may be explained by variations in extinction to cues that differ in their baseline associative strength. For an AY+, AX−, BY− discrimination to be solved, X and B will, presumably, lose more associative strength than A and Y (that are partially reinforced). The associative

strength of B will be less than X from the outset of the discrimination; consequently the individual prediction error for B will be less than the individual prediction error for X. On the one hand, one might expect this to facilitate extinction to X (Rescorla, 2001), resulting in the conditioned responding to AX falling faster than to BY, which we did not observe. On the other hand, however, it is conceivable that the lower prediction error that B will have during discrimination testing will increase attention to B, increasing the overall rate at which associative strength is lost to the BY compound. Taken as a whole, this would result in the discrimination between AY and BY being solved more readily than the discrimination between AY and AX− that is precisely the result observed in Experiment 1. By using a test procedure in which the US can be obtained on each trial with A, B, X, and Y, this possibility is less likely.

Method

Subjects and apparatus. The subjects were 32 experimentally naïve male hooded Lister rats that were maintained in an identical fashion to the subjects in Experiment 1. The apparatus was the same as in Experiment 1.

Procedure. The rats initially received one session of magazine training, and four session of lever press training. The details of these sessions were identical to Experiment 2. All rats were then trained for 16 sessions with an AB+, X+, Y+, P− discrimination, with sessions identical to those used in previous experiments. Following this training, overshadowing was assessed in a test session that was identical to that used in Experiment 1. To reestablish lever pressing, another session of lever training was given in the same manner as before, with a 15-s variable-interval reinforcement schedule.

The final discrimination stage examined rats' ability to use each of the four stimuli as discriminanda for instrumental responding. Sessions in this stage were composed of 32 trials, each 20 s in length, with the same ITIs as used in training and a session length of 71 min. Eight trials were presented with each of A, B, X, and Y, block-randomized so that each successive block of eight trials contained two of each type. During each trial, the two levers were available; they were presented at the same time as the stimulus onset, and retracted at the termination of the stimulus. As for the instrumental conditioning trials of Experiment 2, no lever responses resulted in the delivery of food to the magazine during the first 15 s of each trial; two food pellets were delivered following the first correct response made during the last 5 s of the trial, with a correct response defined as a lever press made on R1 during A or X, and R2 during B or Y. The assignment of levers to R1 and R2 was carried out in the same way as for Experiment 2. Rats received 12 sessions of training during this stage, during which time all instrumental responses made during the initial 15 s of each trial were recorded.

Results

Once again, conditioning in stage 1 proceeded smoothly, and by the final session of conditioning, the mean number of responses per minute made during P, the AB compound, and during the average of X and Y were, respectively, 9.1, 44.3, and 40.3. A one-way ANOVA revealed a difference among these means, $F(2,$

Table 2
Design of Experiment 3

Training	Probe test	Instrumental discrimination
AB+	A−	A: R1+, R2−
X+	B−	B: R1−, R2+
Y+	X−	X: R1+, R2−
P−	Y−	Y: R1−, R2+

Note. A, B, X, Y, and P are all auditory cues; + and − refer to the presence and absence of food, respectively; R1 and R2 are two different instrumental responses. See text for further details.

62) = 63.73, MSE = 186.27. Post hoc tests conducted according to Bonferroni procedure revealed each comparison to be significant, $ts(31) > 2.91$. The number of responses per minute made during the 15-s period before the onset of the CSs was 7.5.

During the test for overshadowing the mean number of responses per minute made during the average of A and B and during the average of X and Y were 26.8 and 32.6, respectively; the difference between these means was significant, $t(31) = 2.27$. The mean number of responses per minute made during the 15-s periods before the test trials during the overshadowing test was 5.2.

During the final test discrimination, the number of responses per minute performed during the first 15 s of the stimuli on the reinforced lever (Correct) and the number of responses per minute performed upon the nonreinforced lever (Incorrect) were averaged across A and B, and across X and Y, and pooled across two-session blocks. The means of these data for the 6 two-session blocks are shown in the top panel of Figure 3. The levers were retracted in the absence of a CS, hence the absence of pre-CS data
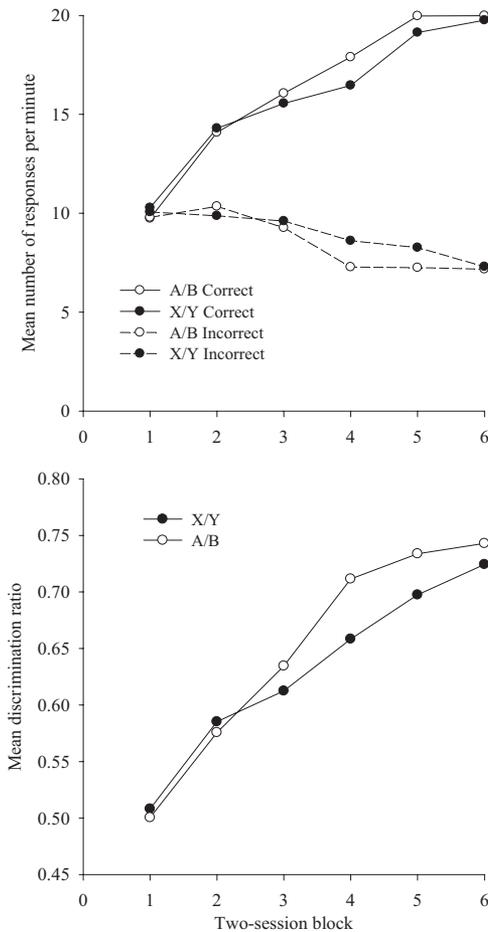
from this figure. By the fourth and fifth session blocks, there is an indication that more correct responses and fewer incorrect responses were made during A and B than during X and Y, implying that the instrumental discrimination was solved better during A and B than during X and Y. To make these data comparable to the data from Experiments 1 and once again to simplify the statistical analysis, discrimination ratios for A/B and X/Y were calculated. These again took the form: p/(p + q). Here, p is now the number of responses per minute made on the correct lever and q is the number of responses per minute made on the incorrect lever. These ratios for the discrimination between A and B and between X and Y are shown in the bottom panel of Figure 3. A two-way ANOVA of individual discrimination ratios with the variables of discrimination (A/B vs. X/Y), and session-block (1 to 6) was conducted. This analysis revealed no effect of discrimination, $F(1, 31) = 2.84$, $MSE = 0.012$, but an effect of session-block, $F(5, 155) = 122.16$, $MSE = 0.004$, and, importantly, an interaction between these variables, $F(5, 155) = 3.48$, $MSE = 0.003$. Simple effects analysis revealed an effect of session-block for both discriminations, $Fs(5, 310) > 58.89$, $MSEs = 0.003$, and a significant difference between the discriminations on session blocks 4 and 5, $Fs(1, 186) > 4.92$, $MSE = 0.004$, remaining $Fs(1, 186) < 1.80$.

Discussion

Rats received trials in which an AB compound, X, and Y were paired with food and P was not. All rats then received probe trials in which A, B, X, and Y were presented in isolation and in extinction. This test revealed conditioned responding to X and Y was more substantial than to A and B, once again demonstrating an overshadowing effect. Cues A, B, X, and Y were then trained as instrumental discriminative stimuli, in which responding on R1 was reinforced during the presence of A and X, and responding during the presence of R2 was reinforced during B and Y. The results showed that stimuli which had been overshadowed (A and B) served as better discriminative stimuli for instrumental conditioning than cues that had previously been conditioned in isolation (X and Y). These results are not consistent with the alternative analysis that was developed for Experiment 2, in which the results of the final test were argued to be a reflection of better learning (and hence more attention) to Y than A. Considering the results of the present experiment, it seems more plausible that performance during the final test of Experiment 2 was a consequence of more effective conditioning with A than with Y.

The results of the current experiment also provide a measure of control over the possibility that CSs that have lower associative strength extinguish faster than CSs that have higher associative strength; a possibility that was raised in the introduction to this experiment and that can be applied to the results of Experiment 1. It may be argued, however, that the results of the experiments reported here simply reflect the fact that stimuli that have lower associative strength have higher associability than stimuli that have higher associative strength. A measure of support for this possibility comes from experiments conducted with pigeons by Dopson, Williams, Esber, and Pearce (2010), who showed that following the solution of an AX+, BX−, CY+, DY− discrimination, the associability of the consistently nonreinforced CSs (B and D) was higher than the associability of the CSs that had been intermittently paired with the US (X and Y). However, associative

Figure 3. Results of the test discrimination from Experiment 3. The top panel shows the mean number of responses performed on the reinforced (correct) and nonreinforced (incorrect) levers during A/B and X/Y across the 6 two-session blocks of the test. The bottom panel shows discrimination ratios calculated from these means.

strength itself cannot be the sole determinant of the associability of CSs. As will be recalled from our description of the experiments conducted by Haselgrove et al. (2010), which used the same stimuli, apparatus and species as in the current experiments, following the solution of an AX+, BY+, X−, Y− discrimination, the associability of A and B (that commanded strong conditioned responding and which, presumably, had high associative strength) was greater than X and Y (that commanded weaker conditioned responding and that, presumably, had lower associative strength). A related result was also reported in autoshaping experiments with pigeons conducted by Dopson, Esber, and Pearce (2010).

An alternative analysis of Experiment 3 might appeal to the effect of blocking. The test session of the current experiment revealed that the rate of responding to A and B was lower than to X and Y, presumably reflecting the fact that A and B have lower associative strength than X and Y. If this is the case then the delivery of the reinforcer upon lever pressing during A and B will have been more surprising than the delivery of the reinforcer upon lever pressing during X and Y. Consequently, standard analyses of learning (e.g., Rescorla & Wagner, 1972) predict that events copresent with A and B (such as lever presses) will be more able to engage in associative learning than will be events copresent with X and Y. One way in which to address this possibility is to note that not only was correct (i.e., reinforced) responding higher during A and B than X and Y, but also that, at least numerically, incorrect (i.e., nonreinforced) responding was lower during A and B than during X and Y. It does not necessarily follow from a model such as that proposed by Rescorla and Wagner that a CS that has higher associative strength will have the capacity to subsequently block both associations between a response and the reinforcer *and* associations between a response and the absence of the reinforcer. However, incorrect performance is not independent from correct performance in this experiment, so it is not entirely clear how fair this rebuttal is. Fortunately, Experiment 2 goes someway to addressing this problem. During instrumental training, A and Y are presented as a compound, and signal reinforcement will be available following performance of R1, but not R2. Under these circumstances, standard theories of learning such as that proposed by Rescorla and Wagner (1972) say that the change in associative strength to A will be equal to the difference between the magnitude of the reinforcer, minus the sum of the associative strengths of all stimuli present on that trial. The same is also true for Y. Consequently, the change in the associative strengths of A and Y during instrumental training should have been equivalent, and thus also should have been their ability to take control of instrumental responding. An analysis based on blocking of reinforcement is therefore incompatible with the results of Experiment 2.

General Discussion

In three appetitive Pavlovian conditioning experiments, rats received trials in which a compound of A and B signaled food and trials in which X and Y were presented alone and which also signaled food. Experiments 1 and 3 both revealed, in subsequent test discriminations, that subdiscriminations that differed on the basis of A and B were more readily solved than subdiscriminations that differed on the basis of X and Y. Experiment 2 showed that when responding on R1, but not R2 was reinforced during a compound of A and Y, A took better control over responding than

Y. Overall, these results are consistent with the idea that a CS conditioned as part of a compound, and that has been overshadowed, has greater associability than a CS that has been conditioned in isolation.

These results are inconsistent with the theory proposed by Mackintosh (1975). According to this theory, weaker responding to A and B than X and Y is a consequence of a loss of associability to A and B, and/or an increase in associability to X and Y. Once this effect is demonstrated, therefore, it follows from Mackintosh's theory that any test that purports to assess associability should reveal higher associability to X and Y than to A and B. To the extent that the tests in the current experiments are appropriate assessments of the associability of A, B, X, and Y, this prediction was not confirmed. Perhaps, then, the current tests are *not* appropriate assessments of the associability of a CS. Although this is a logical possibility, we have strived to use a variety of techniques that, in theory, should be sensitive to variations in the associability of a CS, while at the same time reduced the impact of alternative mechanisms on our results. Furthermore, we, and others have used these techniques in the past to provide evidence *for* Mackintosh's theory (e.g., Duffaud et al., 2007; Haselgrove et al., 2010). If an advocate of Mackintosh's theory were to object to the current techniques for assessing associability, then the same advocate would, presumably, also have to object to these past experiments.

What remains, then, is to try to provide an understanding of these data. A natural starting point when one encounters a result that is diametrically opposite to the prediction of Mackintosh's theory is the theory of learning provided by Pearce and Hall (1980). According to Pearce and Hall, learning is entirely the consequence of variations in the processing of the CS, as the change in associative strength of a CS is given by the product of its associability (α), the magnitude of the US (λ), and a fixed learning rate parameter (S) that is determined by the physical properties of the CS. Associability is increased to CSs that are followed by an unexpected US, and decreased to CSs that are followed by a fully predicted US. This notion is expressed more formally in equation 2.

$$\alpha^{n+1} = |\lambda^n - \Sigma V^n| \qquad (2)$$

Here, α^{n+1} is the associability of the target stimulus on the next trial and $|\lambda^n - \Sigma V^n|$ is an absolute, summed error-term. The summed error-term in the computation for α enables the model to predict effects like overshadowing. However, it also forces the model to make the prediction that the associability of an overshadowed CS will be, at best, equal to a CS conditioned in isolation. This follows because by the end of conditioning the US is fully predicted on AB+ trials, just as it is on, say, X+ trials. Simulations of the Pearce-Hall model for the AB+, X+, Y+ training that we have conducted confirm this and also reveal that before the asymptote of conditioning is reached, the associability of A and B will be, if anything, lower than X and Y, in line with the predictions of Mackintosh's theory. However, at no point during conditioning will the associability of A and B be higher than X and Y. Perhaps, then, by the end of conditioning in the current experiments, the associability of A and B was the same as X and Y, as predicted by the Pearce-Hall model, but the associability of A and B generalized better to the testing conditions than did the associability of X and Y. Although this is a possibility, one might imagine that the similarity of the AB compound to its constituent CSs A

and B will be less than the similarity of X and Y to X and Y themselves (Pearce, 1987). Consequently, one might imagine too that the generalization of associability from the AB compound to A and B in Experiment 3 would be poorer than the generalization of associability from X and Y to X and Y, rather than the converse.

There is a sense, however, in which the results of the current experiments are in keeping with the spirit of the Pearce-Hall model. By the end of AB+ training, the animal can predict that the US will be presented following AB, and under such conditions of certainty the Pearce-Hall model tells us that the associability of the CSs must be reduced. However, it is true to say that there is also considerable *uncertainty* present in the case of AB+ training because, under these circumstances, it is not at all clear from the perspective of the animal whether A or B was responsible for the delivery for the US: A and B are, after all, confounded as predictors of the US. This case is similar in many ways to the philosophical problem of being entirely certain that if two stones are thrown at a window at the same time, then the window will break, but being entirely uncertain as to which stone caused the breakage (Collins, Hall, & Paul, 2004). If the animal were sensitive to this kind of uncertainty it may, in keeping with the spirit of the Pearce-Hall model, pay more attention to A and B than to X and Y that, conditioned by themselves, are unconfounded as predictors of the US. Unfortunately, as the Pearce-Hall model is formally expressed, uncertainty is determined by the difference between the magnitude of the US and the sum of the associative strengths of all CSs present. Perversely, therefore, despite each CS being, by itself, an imperfect predictor of the US (highlighted by the fact that each CS has less than asymptotic associative strength) the Pearce-Hall model tells the animal that associability must be directed toward that CS as if it were a perfectly good predictor of the US. Fortunately, however, a recent variant of the Pearce-Hall model, which has been described by Pearce and Mackintosh (2010), does not make this same assumption.

According to Pearce and Mackintosh (2010), learning progresses according to the summed error-term proposed by Rescorla and Wagner, thus on each trial the associative strength of a CS is updated by the value ($\lambda - \Sigma V$). This value is multiplied by a learning rate parameter, β (determined by the properties of the US), and two learning rate parameters, α and σ, that are modifiable with experience and represent the associability of the CS. α is determined by the rules of associability change proposed by Mackintosh (1975) and these are shown in Equations 1a and 1b. σ is determined by a modification to the Pearce-Hall model. Now, instead of the associability of the CS being equal to an absolute *summed* error-term, it is now equal to an absolute *individual* error-term:

$$\sigma^{n+1} = |\lambda^n - V^n| \qquad (3)$$

Here, V is the associative strength of the CS whose associative strength is being determined. By making this simple change to the way in which the Pearce-Hall component of this "hybrid model" computes associability, the results of the current experiments can be understood. According to the error-term provided by Rescorla and Wagner (1972), by the end of X+ training, the associative strength of X (V_X), will equal λ. Equations 1a and 1b will therefore ensure that α for X will reach its maximum value, as it is the best, indeed the only, predictor of the US on X+ trials. However, Equation 3 will result in σ to X falling to its minimum value by the

end of conditioning: as we have just noted, $V_X = \lambda$; therefore, in this example, the value of $\sigma_X = |\lambda - V_X| = 0$. In contrast, during AB+ training, the error-term provided by Rescorla and Wagner will result in A and B sharing the total amount of associative strength that is supported by the US on that trial. If one assumes that A and B are of equal starting salience, this will result in each acquiring associative strength that is equal to $.5\lambda$. Now, because A and B are equally good predictors of the US, Equations 1a and 1b will ensure that α remains at its starting value.[3] However, in contrast to the case for X that is outlined above, Equation 3 will ensure, for this example, that the value of σ for A and B will always remain positive as $\sigma_{A/B} = |\lambda - V_{A/B}| = .5$.

If one follows the suggestions of Le Pelley (2010; see also Haselgrove et al., 2010) and assumes that the starting values of α and σ are relatively high (0.8), that the minimum obtainable value of $\alpha = .05$, and the minimum obtainable value of $\sigma = 0.3$, then not only can a number of other results that are consistent with an attentional analysis of learning be explained, but so too can the current results. At asymptote, the total associability ($\alpha \times \sigma$) of X following X+ training will be $1 \times 0.3 = 0.3$, and the total associability of A following AB+ training will be $0.8 \times .5 = 0.4$. Consequently, the Pearce-Mackintosh analysis provides, formally, an explanation of why, following AB+, X+, Y+ training, the associability of A and B may be higher than X and Y. At the same time, however, it preserves the nature of the informal analysis developed earlier: Using the Rescorla-Wagner model to calculate associative strength results in A and B sharing the total associative strength that can be supported by the US. Thus, the animal can be certain that the US follows AB (as $V_A + V_B = \lambda$), but he will not be entirely certain why (as $V_A < \lambda$ and $V_B < \lambda$).

If, then, the animal uses the uncertainty of the CS itself to determine its associability, rather than the total uncertainty (or lack of it) from all of the CSs present on that trial, it will be able to subsequently bias its attention to the individual cues of the compound that, themselves, are imperfect predictors of the outcome. Interestingly, the analysis we have just provided generates a way of testing the Pearce-Mackintosh analysis. If A, B, X, and Y entered the experiment with relatively low associability values, rather than high values as we and others have suggested (e.g., Haselgrove et al., 2010; Le Pelley, 2004) then during AB+, X+, Y+ training, α for X and Y will climb to its maximum value, while α for A and B will remain at its low starting value. The consequences of this for the model may be substantial, for now, the total associability of A and B could end up being either equivalent to, or lower than X and Y. This would be a relatively easy prediction to test: all that is required is training before conditioning that either in practice, or in theory reduces the associabilities of A, B, X, and Y. A reasonable candidate for such training would be simple nonreinforced preexposure to the CSs.

---

[3] Note that, although Equation 1b states that the associability of a stimulus will decrease if it is not the sole best predictor of an outcome, the magnitude of this decrease is assumed to be proportional to the difference between the error term associated with the stimulus in question and the error term for all other cues present on that trial. In the case of two stimuli presented in compound, as A and B are here, and that also have equal starting salience, this difference, and therefore the decrease in associability, will be equal to zero. Thus, the salience of these stimuli will remain at its starting value.

Whether this analysis of the associability of overshadowed CSs can be reconciled with the observations of Mackintosh (Mackintosh, 1971, 1976; Mackintosh & Reese, 1979), however, remains to be determined. An explanation for one-trial overshadowing can be provided by both elemental (e.g., Wagner, 2003) and configural (e.g., Pearce, 1987) models of associative learning in terms of generalization decrement. Alternatively, some theorists (e.g., McLaren & Mackintosh, 2000) make the reasonable assumption that learning is continuous, rather than based upon discrete trials. According to this analysis, a single conditioning trial may be thought of as multiple instances in which associative strength is changed. Once these possibilities are acknowledged, then, one-trial overshadowing ceases to be a problem for theories of learning. However, Mackintosh's (1976) observation that overshadowing is attenuated when conditioning is conducted with a compound of CSs that, initially, have equivalent, and high, associability is rather more problematic. Using the summed-error term of the Rescorla-Wagner (1972) model to determine the change in associative strength means that the Pearce-Mackintosh model predicts the presence of overshadowing even when the compound comprises two CS that both have very high associability. It is tempting, of course, to explain away Mackintosh's (1976) result by appealing to the interaction between different levels of conditioned and unconditioned suppression in groups who were conditioned with CSs of different associability. However, this is not very satisfactory, not least because Mackintosh (1976) noted that variations in unconditioned suppression to the different CSs used in his experiment were rather slight. Perhaps a more successful analysis might be achieved by appealing to the possibility, suggested by Mackintosh (1975; see also Le Pelley, Suret, and Beesley, 2009) that α, rather than influencing learning, instead influences performance; thus, conditioned responding to a target CS, say A, is a function of $V_A \times \alpha_A$. According to this analysis, CSs that initially have low or high associability (and which are conditioned in isolation) will, by the end of conditioning, be able to evoke a substantial conditioned response. This follows because these CSs will both acquire asymptotic associative strength, and equation 1a will ensure that α for these two CSs will reach its maximum value. Thus, CSs that start with very different α values and that are conditioned in isolation will eventually acquire equivalent values of α, associative strength, and thus evoke the same level of conditioned responding. Compound conditioning with a pair of CSs that both have, initially, either a high or a low value of α will also result in these CSs acquiring equivalent associative strength (0.5λ). However, equations 1a and 1b will ensure that α for these stimuli remains at, respectively, a high, or a low level. Consequently, following compound conditioning, conditioned responding to the CSs that had a high starting value of α will be substantially stronger than to CSs that had a low starting value of α, a result that will certainly attenuate overshadowing in the case of CSs with high values of α, and, if appropriate limits to performance are assumed, abolish it.

Whatever the virtues, or otherwise, of the analyses provided here, the current results demonstrate a novel feature of Pavlovian conditioning: compound conditioning results in greater associability being acquired by the stimuli that make up that compound than by stimuli that have been conditioned in isolation. This effect is incompatible with traditional theories of conditioning and attention, such as those proposed by Mackintosh (1975) and Pearce and Hall (1980), and suggests a new approach to the determination of stimulus associability is required. We have advocated one solution to this problem here, it remains to be determined if there are others.

References

Beesley, T., & Le Pelley, M. E. (2011). The influence of blocking on overt attention and associability in human learning. *Journal of Experimental Psychology: Animal Behavior Processes, 37,* 114–120.

Blair, C. A. J., & Hall, G. (2003a). Changes in stimulus salience as a result of stimulus preexposure: Evidence from aversive and appetitive testing procedures. *Learning & Behavior, 31,* 185–191.

Blair, C. A. J., & Hall, G. (2003b). Perceptual learning in flavor aversion: Evidence for learned changes in stimulus effectiveness. *Journal of Experimental Psychology: Animal Behavior Processes, 29,* 39–48.

Chamizo, V. D., Aznar-Casanova, J. A., & Artigas, A. A. (2003). Human overshadowing in a virtual pool: Simple guidance is a good competitor against locale learning. *Learning and Motivation, 34,* 262–281.

Collins, J., Hall, N., & Paul, L. A. (2004). *Causation and counterfactuals.* Cambridge, MA: MIT Press.

Dopson, J. C., Esber, G. R., & Pearce, J. M. (2010). Differences in associability of relevant and irrelevant stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 36,* 258–267.

Dopson, J. C., Williams, N. A., Esber, G. R., & Pearce, J. M. (2010). Enhanced attention to stimuli that consistently signal the absence of reinforcement. *Learning and Behavior, 38,* 337–347.

Duffaud, A., Killcross, A. S., & George, D. N. (2007). Optional-shift behaviour in rats: A novel procedure for assessing attentional processes in discrimination learning. *The Quarterly Journal of Experimental Psychology, 60,* 534–542.

Dwyer, D. M., Haselgrove, M., & Jones, P. M. (2011). Cue interactions in flavor preference learning: A configural analysis. *Journal of Experimental Psychology: Animal Behavior Processes, 37,* 41–57.

Dwyer, D. M., & Honey, R. C. (2007). The effects of habituation training on compound conditioning are not reversed by an associative activation treatment. *Journal of Experimental Psychology: Animal Behavior Processes, 33,* 185–190.

Hall, G. (2003). Learned changes in the sensitivity of stimulus representations: Associative and nonassociative mechanisms. *Quarterly Journal of Experimental Psychology, 56*B, 43–55.

Hall, G., & Rodriguez, G. (2009). Factors determining the effects of associative activation on habituation. *Journal of Experimental Psychology: Animal Behavior Processes, 35,* 266–270.

Haselgrove, M., Esber, G. R., Pearce, J. M., & Jones, P. M. (2010). Two kinds of attention in Pavlovian conditioning: Evidence for a hybrid model of learning. *Journal of Experimental Psychology: Animal Behavior Processes, 36,* 456–470.

Holland, P. C. (1999). Overshadowing and blocking as acquisition deficits: No recovery after extinction of overshadowing or blocking cues. *Quarterly Journal of Experimental Psychology, 52*B, 307–333.

James, J. H., & Wagner, A. R. (1980). One-trial overshadowing: Evidence of distributive processing. *Journal of Experimental Psychology: Animal Behavior Processes, 6,* 188–205.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior: Aversive stimulation* (pp. 9–31). Miami, FL: University of Miami Press.

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 830–845.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology, 57*B, 193–243.

Le Pelley, M. E. (2010). The hybrid modelling approach to conditioning. In N. A. Schmajuk (Ed.), *Computational models of conditioning.* Cambridge, UK: Cambridge University Press.

Le Pelley, M. E., Suret, M. B., & Beesley, T. (2009). Learned predictiveness effects in humans: A function of learning, performance or both? *Journal of Experimental Psychology: Animal Behavior Processes, 35,* 312–327.

Mackintosh, N. J. (1971). An analysis of overshadowing and blocking. *Quarterly Journal of Experimental Psychology, 23,* 118–125.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–298.

Mackintosh, N. J. (1976). Overshadowing and stimulus intensity. *Animal Learning and Behavior, 4,* 186–192.

Mackintosh, N. J., & Reese, B. (1979). One-trial overshadowing. *Quarterly Journal of Experimental Psychology, 31,* 519–526.

McLaren, I. P. L., & Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior, 28,* 211–246.

Pavlov, I. P. (1927). *Conditioned reflexes.* London: Oxford University Press.

Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review, 94,* 61–73.

Pearce, J. M., Esber, G. R., George, D. N., & Haselgrove. M. (2008). The nature of discrimination learning in pigeons. *Learning and Behavior, 36,* 188–199.

Pearce, J. M., Graham, M., Good, M. A., Jones, P. M., & McGregor, A. (2006). Potentiation, overshadowing, and blocking of spatial learning based on-the shape of the environment. *Journal of Experimental Psychology-Animal Behavior Processes, 32,* 201–214.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian conditioning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532–552.

Pearce, J. M., & Mackintosh, N. J. (2010). Two theories of attention: A review and a possible integration. In M. E. Le Pelley & C. J. Mitchell (Eds.), *Learning and attention.* Oxford: Oxford University Press.

Rescorla, R. A. (2001). Unequal associative changes when excitors and neutral stimuli are conditioned in compound. *Quarterly Journal of Experimental Psychology, 54*B, 53–68.

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

Revusky, S. (1971). The role of interference in association over a delay. In W. K. Honig & P. H. R. James (Eds.), *Animal memory* (pp. 155–213). New York: Academic Press.

Speers, M. A., Gillan, D. J., & Rescorla, R. A. (1980). Within-compound associations in a variety of compound conditioning procedures. *Learning and Motivation, 11,* 135–149.

Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning.* New York: Academic Press.

Tennant, W. A., & Bitterman, M. E. (1975). Blocking and overshadowing in 2 species of fish. *Journal of Experimental Psychology, 104,* 22–29.

Wagner, A. R. (2003). Context-sensitive elemental theory. *Quarterly Journal of Experimental Psychology Section, 56*B, 7–29.

Wagner, A. R., (1978). Expectancies and the priming of STM. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior.* Hillsdale, NJ: Erlbaum.

Wills, A. J., Lavric, A., Croft, G. S., & Hodgson, T. L. (2007). Predictive learning, prediction errors, and attention: Evidence from event-related potentials and eye tracking. *Journal of Cognitive Neuroscience, 19,* 843–854.