Changes in attention to an irrelevant cue that accompanies a negative attending discrimination

Jemma C. Dopson · Guillem R. Esber · John M. Pearce

Published online: 19 April 2011
© Psychonomic Society, Inc. 2011

Abstract Pigeons were trained in two experiments with negative patterning discriminations that were accompanied by an irrelevant cue. For Experiment 1, the discriminations were of the form AX+ BX+ ABX−, where A and B were relevant, X was irrelevant, and + or − indicate whether or not reinforcement was delivered. The discriminations for Experiment 2 were of the form A+ B+ AX+ BX+ ABX−. A subsequent test phase in both experiments revealed that the associability of A and B, and hence the attention paid to these stimuli, was less than the associability of X. The results are explained with a modified version of a configural theory of associative learning.

Keywords Attention · Negative patterning · Irrelevant cue · Configural theory

There is now good reason to believe that during the course of a relatively simple discrimination, relevant stimuli gain more associability than irrelevant stimuli, an effect that has been equated with the former receiving more attention than the latter (Mackintosh, 1975; Sutherland & Mackintosh, 1971). Many of the experiments from the animal learning literature that support this claim have involved preliminary training with two or more discriminations of the form AX+ BX−, AY+ BY−. In these discriminations, the stimuli A and

B are relevant—as they signal the occurrence and absence of reward, respectively—and are typically from one dimension. The stimuli X and Y are irrelevant, as they are associated with both the occurrence and absence of reward, and are typically from a different dimension. Once these discriminations have been mastered, a second set of discriminations is given, which, for the majority of studies, has involved new values from the dimensions used in the initial training. The new discriminations are typically acquired more readily when stimuli from the previously relevant dimension are the relevant stimuli, than when stimuli from the previously irrelevant dimension are the relevant stimuli. This superior acquisition of discriminations that involve an intradimensional (IDS) rather than an extradimensional (EDS) shift has been demonstrated on a number of occasions (Dias, Robbins, & Roberts, 1996; Mackintosh & Little, 1969; Roberts, Robbins, & Everitt, 1988; Shepp & Eimas, 1964) and can be taken as evidence of more attention being paid to the relevant than to the irrelevant dimension during the initial training (but see Hall, 1991, for an alternative explanation).

A similar IDS–EDS effect has been demonstrated by George and Pearce (1999), but with a more complex initial discrimination than in the foregoing studies. Pigeons received a biconditional discrimination in which food was signalled by two compounds, AP+ BQ+, but not by two other compounds, AQ− BP−. In addition, each of the four compounds was always accompanied by one of two stimuli whose presence was irrelevant to the trial outcome, X and Y. A and B belonged to one dimension, P and Q to a second dimension, and X and Y to a third dimension. A new biconditional discrimination was then introduced, based on the same three dimensions. For an IDS group, the significance of the three dimensions was the same as in the first stage; but for an EDS group, one of the previously

J. C. Dopson (✉) · J. M. Pearce
School of Psychology, Cardiff University,
Cardiff CF10 3AT, Wales
e-mail: dopsonj@cf.ac.uk

J. M. Pearce
e-mail: pearcejm@cardiff.ac.uk

G. R. Esber
Department of Anatomy and Neurobiology,
University of Maryland School of Medicine,
Baltimore, MD, USA

relevant dimensions became irrelevant, and the previously irrelevant dimension became relevant. The new discrimination was acquired more readily by the IDS than by the EDS group. Again, this outcome can be understood if it is accepted that during the test stage subjects paid more attention to stimuli belonging to the previously relevant than to previously irrelevant dimensions.

A different method for demonstrating that more attention is paid to relevant than to irrelevant stimuli has been developed by Pearce, Esber, George, and Haselgrove (2008). Pigeons received a set of discriminations of the form AX+ BX−, CY+ DY−, after which they were given an AX+ AY− CX− discrimination. The AX+ AY− component of the second discrimination is based on the two previously irrelevant stimuli X and Y, whilst the AX+ CX− component is based on the stimuli A and C, which were relevant during initial training. Pearce et al. (2008) found that the AX+ AY− component was acquired more slowly than the AX+ CX− component, and in keeping with the results just described, they attributed the different rates at which the two component discriminations were acquired to more attention being paid to the previously relevant than to previously irrelevant stimuli (see also Dopson, Esber, & Pearce, 2010; Dopson, Williams, Esber, & Pearce, 2010; Haselgrove, Esber, Pearce, & Jones, 2010). Note that for this discrimination, the overall associative strength of each of the three compounds can be regarded as being equivalent at the outset of testing. Therefore, according to theories that assume that the rate of learning on each trial is determined by the combined associative properties of all of the stimuli that are present (e.g., Pearce, 1994; Rescorla & Wagner, 1972), the rates at which the component discriminations are acquired should not differ.

Finally, similar effects have been reported with human participants. For example, Le Pelley and McLaren (2003) conducted a causal reasoning experiment in which participants learned that some cues were relevant and others irrelevant to the occurrence of particular outcomes. When these cues were used to signal a different outcome in a subsequent stage of the experiment, learning progressed more rapidly for the cues that had previously been relevant rather than irrelevant.

Despite the evidence that more attention is paid to relevant than to irrelevant stimuli, a series of experiments by Williams, Mehta, Poworczyk, Oriehel, George and Pearce (2002) suggested that this may not always be the case. These authors found that after rats were trained with the negative patterning discrimination AX+ BX+ ABX−, when X was presented by itself it elicited stronger responding than did a single stimulus that had been repeatedly paired with food throughout the experiment. This result occurred despite X being irrelevant to the solution of the discrimination (see also Williams, Gawel, Reimer, & Mehta, 2005; Williams, Mehta, & Dumont,

2004). If a stimulus elicits a high rate of responding, it might also command a considerable degree of attention, and if this were the case, it would pose a challenge to the generalization that animals pay more attention to relevant than to irrelevant stimuli.

The aim of the two experiments reported here, therefore, was to investigate the changes in attention occurring to relevant and irrelevant stimuli during a negative patterning discrimination. In [Experiment 1](#), a single group of pigeons was trained with the discrimination A+ B+ AB−, in which food was presented whenever A and B occurred independently, but not when they were together. On each trial, one of two stimuli, X and Y, was present, to create an AX+ BX+ ABX− and an AY+ BY+ ABY− discrimination. Thus, A and B can be considered to be relevant to the solution of the discrimination, because their presence is essential if the discrimination is to be solved, whereas X and Y can be considered to be irrelevant. The experiment concluded with an AX+ BX− AY− discrimination, which was of the form of the test discrimination developed by Pearce et al. (2008). If the original training were to result in more attention being paid to A and B than to X and Y, as had been the case in the previous experiments, then the AX+ BX− component of the test discrimination, which is based on A and B, should be acquired more readily than the AX+ AY− component, which is based on X and Y. If, however, under these circumstances more attention were to be paid to the irrelevant cues X and Y, then the AX+ AY− component of the test discrimination should be acquired more readily than the AX+ BX− component.

Williams et al. (2002; see also Williams et al., 2004) concluded that Pearce's (1987, 1994, 2002) configural theory, rather than an elemental theory of learning, was able to provide the best explanation for the results of their experiments. However, this theory does not include a mechanism for producing changes in attention to stimuli. In view of this shortcoming, Pearce, George, and Redhead (1998) devised a modification to Pearce's (1994) configural theory that allows it to make predictions concerning the changes in attention that occur during both complex (e.g., George & Pearce, 1999) and simple (e.g., Mackintosh & Little, 1969) discriminations. The modification remains relatively untested, and hence an additional benefit of our experiments is that they will provide an opportunity to test this revision to configural theory. In order to avoid undue repetition, a description of the revised theory and of the predictions it makes concerning the experiments is postponed until the [General Discussion](#).

Experiment 1

A single group of pigeons was first trained with the negative patterning discriminations described above and

summarised in the upper part of Table 1. The stimuli were presented on a television screen behind a clear response key as the conditioned stimuli (CS) for autoshaping. Each letter represents a single circle containing a unique pattern. If subjects were to only receive training with AX+ BX+ ABX– and AY+ BY+ ABY–, they might solve the discrimination by referring to the number of circles on the screen. In order to forestall this strategy, the experiment included a positive patterning discrimination with a common element on every trial, CZ– DZ– CDZ+, so that the number of circles on the screen was no longer informative about the outcome of a trial. Towards the end of the initial training, and prior to the test discrimination shown in Table 1, probe test trials were conducted with A, B, X, and Y, in order to compare the rates of responding elicited by the relevant and irrelevant stimuli. In view of the results of Williams et al. (2002), it was expected that responding during the irrelevant stimuli would be particularly high. In order to examine the associability of the irrelevant stimuli, the experiment concluded with the test discrimination described above and summarised in Table 1.

Method

Subjects The subjects were 14 experimentally naïve adult homing pigeons. The pigeons were housed in pairs and had free access to water and grit in their home cages. They were maintained at 80% of their free-feeding weights by being fed a restricted amount of food after each experimental session. They were maintained in a lightproof room in which the lights were on for 14.5 h each day.

Apparatus The experimental apparatus consisted of eight pigeon test chambers (30.0 × 33.0 × 35.0 cm). Each contained an 8.3-cm-high × 6.3-cm-wide clear acrylic panel, which was hinged at the top. Pecks on the panel were detected by a reed relay that was operated whenever a magnet attached to its lower edge was displaced a distance greater than 1 mm. The midpoint of the panel was 24 cm above the floor of the chamber. A colour thin film transistor TV with a 15.5 × 8.7 cm screen was located 2.0 cm behind the acrylic panel. Food was delivered by operating a grain

feeder (Colbourn Instruments, Lehigh Valley, PA) with an opening that measured 4.6 × 5.4 cm located in the same wall as the response key. The midpoint of the opening was 9.0 cm above the chamber floor and 7.0 cm to the left of the midline of the wall. The feeder was illuminated whenever grain was made available. The chambers were permanently lit during all experimental sessions by a 2.8-W bulb, operated at 24 V, located in the centre of the ceiling of the chamber. The control of events, recording of responses, and generation of stimuli on the TV screens took place with a PC computer (Research Machines, Abingdon, U.K.) running Windows XP. The computer was programmed in Visual Basic, and the interface with the experimental apparatus was controlled by Whisker software (Campden Instruments Ltd, Loughborough, U.K.).

Stimuli The stimuli were presented on the TV screen, which was otherwise black. Each of the letters in the compounds presented in Table 1 was represented by one circle, which was 2.4 cm in diameter. Therefore, on AX, BX, AY, BY, CZ, and DZ trials there were two circles on the screen, and on ABX, ABY, and CDZ trials there were three circles on the screen. The circles could be in one of three positions: with the midpoint of the circle 1.2 cm left of the midpoint of the screen (left), with the midpoint of the circle 1.2 cm right of the midpoint of the screen (right), or with the midpoint of the circle at 1.2 cm above the midpoint of the screen (top). When three circles were on the screen, there would be a circle in each of these positions. When two circles were on the screen, there would be a circle in two of the positions. Trial types were presented in three different ways to ensure that each element appeared in each of the three positions. Thus, the three different configurations of the ABX trial type, for example, were A on the left, B at the top, and X on the right; B on the left, X at the top, and A on the right; and X on the left, A at the top, and B on the right. The different configurations of AX were A on the left and X on the right; X at the top and A on the right; and X on the left and A at the top. Examples of these configurations are shown in Fig. 1.

Each of the circles was filled with one of six patterns. Four of the patterns were created from three black bars,

Table 1 Summary of the different trials that were used for Stage 1 and the test stage of Experiments 1 and 2

Stage 1				Test		
Experiment 1						
AX+	BX+	2ABX–		AX+	AY–	BX–
AY+	BY+	2ABY–				
CZ–	DZ–	2CDZ+				
Experiment 2						
A+	B+	AX+	BX+	2ABX–	AX+	AY–
C+	D+	CY+	DY+	2CDY–		CX–

Fig. 1 Examples of the stimuli from Experiment 1. The three configurations for an AX+ trial are shown in the left-hand column, and the three configurations for an ABX- trial are shown in the right-hand column

each 3 mm wide, separated by white bars of the same width. For each pattern, the orientation of the bars with respect to the top of the monitor was either 0°, 45°, 90°, or –45°. Three black-and-white patterns were used in addition to these four stimuli. One of these consisted of three white circular bands, each with a width of 2 mm. The outermost band formed a circle with diameter 2.4 cm. Each band was separated by black circular bands of the same width. In the centre of the stimulus was a black circle (diameter 4 mm). A “spotted” pattern consisted of a white circle containing seven black circles, each with diameter 0.64 cm, and the final pattern was a black five-point star within a white circle with diameter of 2.4 cm. The points of the star touched the edges of the circle, with the top point being directly above the centre.

For half of the birds, A and B represented the patterns with lines of orientation 90° and 0°, respectively, and X and Y represented the concentric circle and spotted patterns. For the other half, A and B were the concentric circle and spotted patterns, and X and Y were the 90° and 0° patterns. For all subjects, C, D, and Z, the patterns used in the control discrimination, represented the lines oriented 45°, the star, and the lines oriented –45°, respectively.

Procedure The subjects first received 8 sessions in which they were trained to eat food whenever it was presented by the hopper. They were then given five sessions of autoshaping in which a white circle with a diameter of 2.4 cm was presented in the middle of the TV screen for 10 s. There were 45 trials in a session, the mean interval between the start of separate trials was 60 s (range, 40–80 s), and food was made available in the hopper for 4 s whenever the white circle was removed from the TV screen.

Session 1 of Stage 1 began on the day after the final session of autoshaping with the white circle. During each of

the 40 sessions of training, subjects received discrimination training with the 9 trial types given in the upper part of Table 1. In each session there were 72 trials. During a single session, trial types AX, AY, BX, BY, CZ, and DZ were each presented 6 times (with each of the three configurations of stimuli appearing twice). In order to equate exposure to reinforced and nonreinforced stimuli, trial types ABX, ABY, and CDZ were each presented 12 times during each session (with each of the three configurations of stimuli appearing four times). Each stimulus was presented for 10 s, and the mean intertrial interval was 60 s (range = 40–80 s). On reinforced trials, food was presented in the same manner as during the autoshaping stage, and on nonreinforced trials, no event followed the removal of the stimuli from the screen. The trial types were presented in a random order, with the constraint that no more than three reinforced or nonreinforced trials occurred in succession.

During Session 36 of Stage 1, all subjects received 72 training trials, as in each of the other sessions of this stage. These trials were intermixed with 8 nonreinforced trials with only one circle (diameter 2.4 cm) in the centre of the screen. This circle contained A, B, X, or Y, with each of these four trial types being presented twice.

On the day following Session 40 of training, subjects received the first session of the test discrimination. There were 10 sessions in this stage, during which all animals received a discrimination involving three trial types that had been reinforced during training: AX+ AY– BX–. There were 40 trials in each session, consisting of 20 reinforced presentations of AX, 10 nonreinforced presentations of AY, and 10 nonreinforced presentations of BX. The procedural details that have been omitted were the same as for the previous stage.

Results

A Type I error rate of $p < .05$ was adopted for all of the statistical tests. Tests of simple main effects were conducted with a pooled error term (Kirk, 1968). The rate of pecking at the Perspex panel in front of the TV screen was recorded on every trial. The mean rate of responding during the nine trial types presented in Stage 1 was calculated for each bird for every five-session block. The means of these data are presented separately in Fig. 2 for the three discriminations (AX+ BX+ ABX–, AY+ BY+ ABY–, and CZ– DZ– CDZ+). From this figure it is evident that the negative patterning discriminations were acquired at similar rates. Although the positive patterning discrimination was not acquired to the same extent, responding during the reinforced trial types was more rapid than during the nonreinforced trial types by the end of Stage 1.

Fig. 2 The group mean rates of responding during the AX+, BX+, and ABX- trials (left-hand panel); AY+, BY+, and ABY- trials (centre panel); and CZ-, DZ-, and CDZ+ trials (right-hand panel) in five-session blocks of Stage 1 of Experiment 1

The mean response rates across the final block of five sessions of Stage 1 for each of the three trial types in each negative patterning discrimination were compared using *t* tests. Bonferroni-corrected pairwise comparisons revealed that for the AX+ BX+ ABX- discrimination, the rates of responding during the trials with AX and BX did not differ, $t(13) = 0.23$, but that the response rate during each of these trial types differed significantly from that during ABX-, $t_s(13) > 5.16$. Similarly, for the AY+ BY+ ABY- discrimination, the rates of responding during AY and BY did not differ, $t(13) = 0.15$, but the response rate during each of these trial types differed significantly from that during ABY, $t_s(13) > 5.99$. The analyses for the CZ- DZ- CDZ+ discrimination revealed that the rate of responding during CZ trials across the final block of five sessions was significantly slower than that during DZ trials, $t(13) = 3.37$. Responding during CDZ trial types was significantly faster than during either of these trial types, $t_s(13) > 4.20$.

The combined mean rate of responding during the test trials with A and B when they were presented individually was 141.7 responses per min. The mean rate of responding during the equivalent trials with X and Y was also high, at 196.6 responses per minute. This difference was not statistically significant, $t(13) = 1.54$. In addition, during the same session, the rate of responding during X was not significantly different from the mean rate during AX and BX, and the rate of responding during Y was not significantly different from the mean rate during AY and BY, $t_s(13) < 0.52$.

The mean rates of responding during presentations of the reinforced compound AX and nonreinforced compounds AY and BX are shown for each session of the test stage in Fig. 3. Inspection of this figure reveals that responding to AX remained at a high rate throughout the test stage, and responding to AY decreased more rapidly than responding to BX. In order to compare performance

during the two nonreinforced compounds, a two-way ANOVA of individual mean response rates with the factors of Trial Type (AY and BX) and Session was conducted. This analysis revealed significant effects of trial type $F(1, 13) = 12.26$, $MSE = 11,593.23$, and of session, $F(9, 117) = 26.11$, $MSE = 1,684.14$, and the interaction was also significant, $F(9, 117) = 4.95$, $MSE = 1,119.88$. Simple-effects tests revealed that responding to AY was at a rate lower than BX's on Sessions 4, 5, 6, 7, 8, 9, and 10, $F_s(1, 130) > 7.78$, $MSE = 2,167.21$.

Discussion

The significantly slower rate of responding during AY trials than during BX trials indicates that the AX+ AY-

Fig. 3 The group mean rates of responding during the reinforced compound, AX+, and the nonreinforced compounds, AY- and BX-, for the 10 sessions of the test discrimination of Experiment 1

component of the test discrimination was acquired more readily than the AX+ BX– component. The implication of this finding is that during the test stage of the experiment, more attention was paid to the previously irrelevant stimuli, X and Y, on which the former component of the test discrimination was based, than to the previously relevant stimuli, A and B, on which the latter component was based. This outcome stands in marked contrast to many experiments in which researchers have found, with different methods of training, that more attention was paid to relevant than to irrelevant stimuli (e.g., Mackintosh & Little, 1969). The present results were not, however, entirely unexpected. Our method of training was similar to that in a series of experiments by Williams et al. (2002; see also Williams et al., 2005; Williams et al., 2004), in which it was found that the common element of an AX+ BX+ ABX– discrimination elicited an above-asymptotic rate of responding when presented by itself. If a stimulus elicits a high rate of responding, it is not unreasonable to expect it to be paid considerable attention. The test trials with X and Y in the present experiment revealed that they too elicited a high rate of responding by themselves. This rate was numerically greater than the rate to A and B, but not significantly so. On this basis, therefore, the results of the test stage should not be too surprising.

Experiment 2

An inspection of the upper part of Table 1 indicates that during one complete cycle of all the training trials, subjects would have experienced each of X and Y four times, and each of A and B six times. Repeated exposure to a stimulus will result in latent inhibition or, in other words, a loss in its associability (e.g., Hall, 1991), and it is possible that the different amounts of exposure during the training stage resulted in the associability of X and Y being higher than that of A and B at the outset of the test stage. Experiment 2 was conducted with this possible explanation for the outcome of Experiment 1 in mind.

The design of Experiment 2 was based on that of Experiment 1. Subjects again received an AX+ BX+ ABX– discrimination, but they also received additional reinforced trials with A and B by themselves (see the lower part of Table 1). Williams et al. (2005) have shown that the addition of the reinforced trials with A and B does not prevent X from eliciting a high rate of responding when it is presented by itself once the discrimination has been mastered. It was anticipated, therefore, that despite the change in design, the associability of the irrelevant X would again be high at the end of the initial discrimination. In order to provide a suitable set of stimuli for the test

stage, the lower part of Table 1 shows that the initial training also included a C+ D+ CY+ DY+ CDY– discrimination, where C and D were relevant and Y was irrelevant. Again, an AX+ AY– CX– discrimination was given in the subsequent test stage, and it was anticipated that, in keeping with the results of Experiment 1, the AX+ AY– component of the discrimination would be acquired more readily than the AX+ CX– component. Because the design of this experiment ensured that A and C were presented as often as X and Y during Stage 1, it would be difficult to attribute this result to latent inhibition being more disruptive for learning about A and C than for learning about X and Y.

A positive patterning discrimination was included in Experiment 1 in order to prevent subjects from only using the number of circles to solve the discrimination. For similar reasons, a preliminary study was conducted using the design shown in the lower part of Table 1, but with an additional positive patterning discrimination based on the one in Experiment 1. Initial training revealed that it was particularly difficult for pigeons to master the discriminations, and for the present experiment it was therefore decided to omit these additional trials. Although the number of circles on the screen might be used as a cue for solving the discriminations in Stage 1, it could not be responsible for any differences observed during the test stage, as there were only two circles on the screen during each test trial.

Method

Subjects and apparatus The subjects were 15 experimentally naïve adult homing pigeons. The pigeons were housed in pairs and had free access to water and grit in their home cages. They were fed and maintained in the same manner as in Experiment 1 and tested in the same apparatus used for Experiment 1.

Stimuli The stimuli were presented in the same way as in Experiment 1. When the trial type consisted of just one stimulus, one circle appeared on the screen. Each of these trial types was presented in three different ways to ensure that the circle appeared in each of the three positions described for Experiment 1. The AX, BX, CY, and DY trials each appeared in the three different configurations described for the two-element compounds in Experiment 1. In this experiment, the six possible configurations of ABX and CDY were used. The same patterns used in Experiment 1 were used for Experiment 2, as well as an additional pattern, which comprised a black “S” in the centre of a white circle.

For 8 of the birds, the stimuli presented during the test discrimination, A, C, X, and Y, were circles containing parallel black and white lines oriented at 90° or 0°, a circle containing a set of concentric circles, and the white circle

containing the small black circles, respectively. For the remaining 7 birds, the assignment of the stimuli to A, C, X, and Y was reversed. The remaining stimuli, which were black and white lines oriented at $\pm 45^\circ$, the “S,” and the star, were assigned in such a way that each training compound of two stimuli comprised one circle with and one circle without black and white lines.

Procedure Magazine training and autoshaping were conducted in the same manner as for Experiment 1. Session 1 of Stage 1 began on the day after the final session of autoshaping. During each of 50 sessions, subjects received discrimination training with the 10 trial types given in Table 1. In each session there were 36 trials. During a single session, trial types A, B, C, D, AX, BX, CY, and DY were each presented three times, and ABX and CDY were each presented six times (each of the different configurations of the trial types appeared once). Procedural details that have been omitted were the same as for Experiment 1.

During Session 48 of Stage 1, all subjects received training trials intermixed with two nonreinforced trials on which X was presented in the centre of the screen, and two nonreinforced trials on which Y was presented.

On the day following Session 50 of training, subjects received the first session of the test discrimination. There were 10 sessions, during which all animals received a discrimination involving one of the trial types that had been reinforced during training, AX, and two novel compounds that consisted of one element that had been relevant and another that had been irrelevant during Stage 1, AY and CX. There were 40 trials in each session of the test discrimination, consisting of 20 reinforced presentations of AX, 10 nonreinforced presentations of

AY, and 10 nonreinforced presentations of CX. The procedural details that have been omitted were the same as for the previous stage.

Results

The group mean rate of responding during the 10 trial types presented in Stage 1 was calculated for each bird for every five-session block. Figure 4 shows these data separately for the two discriminations. With continued training, responding for both discriminations during the nonreinforced compounds became less rapid than during the remaining reinforced trials. In order to confirm that the two discriminations had been mastered by the end of Stage 1, the mean rate of responding during each nonreinforced compound for the final five sessions of training was compared with the equivalent rate for each of the different types of trial with the components of the compound. The rate of responding during ABX was significantly slower than during A, B, AX, or BX, $t_s(14) > 4.59$. Likewise, the rate of responding during CDY was significantly slower than during C, D, CY, or DY, $t_s(14) > 4.45$.

The mean rate of responding during the test trials with the two irrelevant stimuli, X and Y, was 189.2 responses per min. This rate was similar to that recorded during the training trials in the same session on which the other two stimuli used for the test stage, A and C, were presented alone; A and C elicited a mean of 180.9 responses per minute. This difference between the rates of responding during the relevant and irrelevant stimuli was not statistically significant, $t(14) = 1.16$.

The mean rates of responding during presentations of the reinforced compound AX and the nonreinforced com-

Fig. 4 The group mean rates of responding during the A+, B+, AX+, BX+, and ABX– trials (left-hand panel) and the C+, D+, CY+, DY+, and CDY– trials (right-hand panel) in five-session blocks of Stage 1 of Experiment 2

pounds AY and CX are shown for each session of the test stage in Fig. 5. Inspection of this figure reveals that, as in Experiment 1, responding to AX remained at a high rate throughout the test stage, and responding to AY decreased more rapidly than responding to CX. The difference between the response rates to AY and CX was investigated by conducting a two-way ANOVA with the factors of Trial Type (AY and CX) and Session. This analysis revealed a significant effect of session, $F(9, 126) = 14.08$, $MSE = 1,648.99$. The Trial Type $\times$ Session interaction was also significant, $F(9, 126) = 2.39$, $MSE = 1,023.83$, and simple-effects tests revealed that responding to AY was at a lower rate than responding to CX on Sessions 5, 6, 7, and 8, $F_s(1, 140) > 4.90$, $MSE = 1,983.64$.

Discussion

Despite the differences in design, the results from the present experiment were similar to those of Experiment 1. In both cases, a component of the test discrimination that was based on previously irrelevant stimuli was acquired more readily than a component based on previously relevant stimuli. In contrast to Experiment 1, the amounts of exposure to the relevant and irrelevant stimuli were the same in the present study, which makes it unlikely that the outcome was a consequence of latent inhibition having a greater disruptive influence on learning about the relevant than about the irrelevant stimuli.

The absence of a positive patterning discrimination in Stage 1 is unfortunate, because subjects may have made

use of the number of circles on the screen as a cue for solving the discriminations in that stage. It is thus possible that the high rate of responding during the test trials with X and Y towards the end of Stage 1 was a consequence of one circle on the screen having previously served as a cue for food. On the other hand, since every trial during the test stage involved two circles, some factor other than the number of circles on the screen must have been responsible for the different rates at which responding during AY and CX decreased during this stage of the experiment. Throughout this article, we have assumed this factor is the amount of attention paid to the previously relevant and irrelevant stimuli. The main purpose of the discussion that follows is to consider different explanations for how the original training given in Experiments 1 and 2 might result in more attention being paid to the irrelevant than to the relevant stimuli involved. It is also necessary to consider how these explanations might be applied to findings that suggest that during other kinds of discrimination learning animals pay more attention to relevant than to irrelevant stimuli.

General discussion

In the two reported experiments, pigeons were trained with a negative patterning discrimination for which a common element X was present on every trial: AX+ BX+ ABX−.

Both experiments showed that as a result of this training, the amount of attention paid to this common element was high, relative to A or B. It thus appears that this particular discrimination results in an irrelevant stimulus acquiring the properties that would normally be expected of a stimulus that is relevant to its solution.

Williams et al. (2002; see also Williams et al., 2005; Williams et al., 2004) trained rats with a similar negative patterning discrimination, and subsequent test trials revealed an above-asymptotic strength of conditioned responding, or *superconditioning*, to X. In both experiments reported here, although responding to the irrelevant stimuli alone was strong, there was no indication of the superconditioning effect reported by Williams et al. (2002). That is, the rate of responding during the test trials with X and Y was equivalent to, but not faster than, that recorded during relevant stimuli that had been consistently reinforced. This failure to replicate the findings of Williams et al. (2002; Williams et al., 2005) may have occurred because a performance ceiling at the time of testing masked any superconditioning with X and Y. Alternatively, the feature negative discriminations were acquired slowly, and the associative strength of the irrelevant stimuli may not have had time to reach an asymptotic, abnormally high level when testing took place.

Fig. 5 The group mean rates of responding during the reinforced compound, AX+, and the nonreinforced compounds, AY− and CX−, for the 10 sessions of the test discrimination of Experiment 2

The principal purpose of the experiments was not to test whether the training in Stage 1 resulted in superconditioning with X and Y, but to examine how this treatment affected the attention paid to these stimuli. We now consider the implications for different theories of attention of our finding that attention to X and Y was high at the outset of the test phase. According to Mackintosh (1975), attention to a CS will increase if it is a better predictor of the trial outcome than any other stimulus, and decline if this is not the case. The rate of conditioning to a CS is then determined by the attention paid to that CS and by the discrepancy between its current associative strength and the asymptote of conditioning, λ . Equation 1 shows how these two factors combine to determine the change in associative strength to CS A, V_A , on a single trial, where α_A is the amount of attention paid to A and β is a learning rate parameter whose value is set by properties of the unconditioned stimulus (US):

$$\Delta V_A = \alpha_A \cdot \beta \cdot (\lambda - V_A). \quad (1)$$

A limitation of Eq. 1 is that it fails to provide a satisfactory account of how certain discriminations are solved. Le Pelley (2004, pp.202–203) has pointed out that it is difficult for Eq. 1 to explain the solution of a feature negative discrimination, A+ AX–, and it is not clear how the equation should be applied to the discriminations used in Experiments 1 and 2 above. In order to broaden the applicability of the principles of attention proposed by Mackintosh (1975), therefore, a number of theories have combined them with the principles of association formation advocated by Rescorla and Wagner (1972). To appreciate the significance of the present results for this class of theory, we shall focus on the most recent version, put forward by Pearce and Mackintosh (2010). For related theories, see Buhusi and Schmajuk (1996), Le Pelley (2004), Moore and Stickney (1985), and Schmajuk and Moore (1985). According to Pearce and Mackintosh, the change in associative strength to CS A on any trial is given by Eq. 2, in which V_T represents the sum of the associative strengths of all the stimuli present on a trial:

$$\Delta V_A = \alpha_A \cdot \sigma_A \cdot \beta \cdot (\lambda - V_T). \quad (2)$$

Equation 2 has followed Le Pelley (2004) by incorporating a learning rate parameter σ_A , the value of which is set according to principles based on the Pearce and Hall (1980) theory. Although the theory of Pearce and Mackintosh (2010) is able to account for a wider range of phenomena than the theories on which it is based, it encounters a problem when applied to the A+ B+ AX+ BX+ ABX– discrimination of Experiment 2. As it has been presented thus far, the theory predicts that responding during ABX throughout training should be stronger than during any other

trial, whereas it is quite evident that as training progressed the opposite outcome was found.¹

The foregoing prediction can be avoided if stimulus compounds create configural cues that enter into associations in the same way as conventional cues (Wagner & Rescorla, 1972). The problem now is that a number of decisions must be made before precise predictions can be derived from the theory. One decision concerns the number of configural cues that will be present on each type of trial. Another decision relates to whether the saliences of the configural cues and the experimental stimuli should be the same or different. Decisions must then be made about the values that should be ascribed to the parameters that determine the rate at which α and σ vary during the course of training. At present, there is insufficient evidence to allow any of these decisions to be made with confidence, which makes it difficult to derive precise predictions from Pearce and Mackintosh's (2010) proposals concerning the above experiments. Rather than speculate on how these decisions might be made, we now turn to an alternative account of how discriminations are solved. We then explore how this alternative can be modified to explain the results from Experiments 1 and 2.

Williams et al. (2002; Williams et al., 2004) referred to the configural theory of Pearce (1987, 1994, 2002) in order to explain the effects of both an AX+ BX+ ABX– and an A+ B+ AX+ BX+ ABX– discrimination. According to this theory, changes in associative strength on any trial take place to configurations of stimuli, which can be represented as hidden units in a connectionist network. Figure 6 shows the configural units that will be formed during an AX+ BX+ ABX– discrimination. For the sake of simplicity, the only connections shown are those that will be activated on an AX trial. The letters in boxes on the left represent input units activated by individual stimuli, and the boxes on the right are output units that represent the presence and omission of the US and are responsible for generating a conditioned response (CR). A configural unit will be maximally activated, via the solid lines, when the pattern to which it corresponds is presented to the input layer. If the pattern at the input layer should bear some similarity to the pattern represented by a configural unit, then the unit will be activated to a degree that is determined by the similarity of the two patterns. Thus, on an AX trial, the configural unit for AX will be activated to its maximal level of 1, the unit for BX will be activated to .25, and the unit for ABX will be activated to .66 (these values were

¹ The theory of Pearce and Mackintosh (2010) does not run into this problem with Experiment 1, where it predicts that the discrimination will be solved when the associative strength of the irrelevant cues is 2λ and that of the relevant cues is $-\lambda$. In view of the challenges posed by the results of Experiment 2 to the theory of Pearce and Mackintosh, we have not pursued its application to Experiment 1.

Fig. 6 The connectionist network that will be formed, according to Pearce (1994), during an AX+ BX+ ABX– discrimination. The only connections shown are those activated on an AX trial. The dashed lines indicate feedback links proposed by Pearce et al. (1998). Circles depict configural units, small boxes depict input units, and large boxes depict output units

calculated from an equation presented in Pearce, 1994). The reinforced trials will result in the configural units for AX and BX entering into excitatory associations with the US, E_{AX} and E_{BX} , whereas the nonreinforced trials will result in ABX entering into an inhibitory association with the US, via a no-US unit, $-E_{ABX}$. The level of activation of a configural unit, multiplied by its associative strength, will determine its contribution to the activation of the output unit to which it is connected. At asymptote, AX and BX will elicit a CR, and there will be no response during ABX. Computer simulations based on equations proposed by Pearce (1994) reveal that at asymptote, the value of both E_{AX} and E_{BX} will be 2.77λ and the value of E_{ABX} will be -3.69λ , respectively. It then follows that when X is presented by itself, it will elicit a CR of magnitude 1.53λ . For the more complex A+ B+ AX+ BX+ ABX– discrimination, this value is reduced to 1.20λ . Williams et al. (2002; Williams et al., 2004) provide a more detailed discussion of the application of the theory of Pearce (1994) to the two discriminations.

In order to allow the network shown in Fig. 6 to enable changes in attention to take place, Pearce et al. (1998) added the feedback links depicted by the dashed lines. During a conditioning trial, any activated configural unit that predicted the correct outcome (reward or nonreward) was assumed to feed a positive signal back to all of the active inputs to which it was connected. The magnitude of this signal was determined by the degree to which the configural unit was activated. A similar process was assumed to occur if the

configural unit predicted the wrong outcome, but the feedback signal would be negative. The purpose of this feedback was to modify the effectiveness of the input unit for activating the configural unit, with net positive feedback increasing its effectiveness and net negative feedback having the opposite effect. Such a change in the effectiveness of input units can be regarded as being equivalent to a change in attention, since discriminations between compounds that excite highly effective input units will be easier to solve than discriminations in which the input units excited are only weakly effective.

The application of the foregoing principles to an AX+ BX+ ABX– discrimination leads to the prediction that more attention will be paid to X than to A or B. On AX and BX trials, X will receive positive feedback from both AX and BX configural units, whereas on these trial types A and B will only receive positive feedback from one configural unit (AX or BX, respectively). Therefore, the net effect will be more positive feedback to X than to either A or B. However, when additional trials on which X is not presented are included (AY+, BY+, and ABY– trials in Exp. 1, or A+ and B+ trials in Exp. 2), more positive feedback will accrue to A and B than in the standard negative patterning discrimination. As a consequence, the model predicts that following these more complex discriminations, more attention will be paid to the stimuli that were originally relevant than to those that were originally irrelevant during the test stage. As we have seen, however, the opposite outcome was found, and an alternative to the modification proposed by Pearce et al. (1998) must be sought.

There is, of course, no compelling reason why feedback to the input units should be related to the level of activation of the configural units to which they are connected. From the point of view of directing attention to stimuli that are important, the same network shown in Fig. 6 could send back to an input unit a signal directly related to the associative strengths of the configural units to which it is connected, rather than related to the degree to which those configural units are activated. The sign of this feedback would again be positive, if the configural unit predicted the correct outcome of the trial, and negative, if it predicted the incorrect outcome. Thus, on any trial, the feedback will be positive either if the associative strength of the unit is positive and a US is presented, or if the associative strength of the unit is negative and a US is not presented.

The application of this idea to the A+ B+ AX+ BX+ ABX– discrimination of Experiment 2 can be understood with reference to Table 2. The top row of this table depicts the configural units that will be formed during the discrimination, and the values below each configural unit represent its asymptotic associative strength as predicted by equations described in Pearce (1994). The asymptote of conditioning, λ , was set at 1. The top row of the upper panel of the table shows the feedback that will be directed to Input Unit A on an A+ trial from each of the configural units to which A is connected. The remainder of this panel of the table shows the feedback to Input Unit A that will

accrue from each of the remaining trials of the discrimination, and that the strength of the resultant signal from all of the feedback will be -0.5 . A similar summary for the feedback that will be received by Input Unit X is presented in the lower panel of the table, from which it can be seen that the resultant signal will be 1.0 . According to this analysis, therefore, X will receive more feedback than A, and will thus effectively be paid more attention. Informally, the explanation for this prediction is that during Stage 1, X belongs to two configurations paired with food (AX and BX) that each gain considerable excitatory strength, whereas A belongs to one configuration that gains high excitatory strength (AX) and one that gains moderate excitatory strength (A). There is thus the opportunity for more positive feedback to accrue to the input unit for X than to the input unit for A.

A similar prediction is made concerning Experiment 1, even though the analysis is slightly more complex. The top two rows of Table 3 show the six configurations that were used for Stage 1 of this experiment, together with their asymptotic associative strengths as predicted by the theory of Pearce (1994). The rest of this table shows how the feedback to the input units of A and X will be determined by the various training trials. The total feedback for the entire discrimination is then predicted to be greater to the input unit for X (1.6) than to the input unit for A (0.0). The informal explanation for this prediction is that X belongs to one nonreinforced configuration (ABX) and two reinforced

Table 2 Calculation of the feedback to Input Units A and X according to the modification to the theory of Pearce (1994) for the A+ B+ AX+ BX+ ABX– discrimination of Experiment 2

	Configural Unit				
	A+	B+	AX+	BX+	ABX–
Input Unit A	1.1	1.1	2.6	2.6	–4.2
Trial					Feedback per Trial
A+	1.1		2.6		–4.2
B+					
AX+	1.1		2.6		–4.2
BX+					
ABX–	–1.1		–2.6		4.2
Total feedback to Input Unit A for entire discrimination:	–0.5				
Input Unit X					
Trial					Feedback per Trial
A+					
B+					
AX+			2.6	2.6	–4.2
BX+			2.6	2.6	–4.2
ABX–			–2.6	–2.6	4.2
Total feedback to Input Unit X for entire discrimination:	1.0				

Table 3 Calculation of the feedback to Input Units A and X according to the modification to the theory of Pearce (1994) for the AX+ BX+ ABX– AY+ BY+ ABY– discrimination of Experiment 1

		Configural Unit						
		AX+	BX+	ABX−	AY+	BY+	ABY−	
Input Unit A		1.9	1.9	−2.2	1.9	1.9	−2.2	
Trial							Feedback per Trial	
AX+	1.9			−2.2	1.9		−2.2	−0.6
BX+								
ABX−	−1.9			2.2	−1.9		2.2	0.6
AY+	1.9			−2.2	1.9		−2.2	−0.6
BY+								
ABY−	−1.9			2.2	−1.9		2.2	0.6
Total feedback to Input Unit A for entire discrimination:								0.0
Input Unit X								
Trial								Feedback per Trial
AX+	1.9	1.9		−2.2				1.6
BX+	1.9	1.9		−2.2				1.6
ABX−	−1.9	−1.9		2.2				−1.6
AY+								
BY+								
ABY−								
Total feedback to Input Unit X for entire discrimination:								1.6

configurations (AX and BX), whereas A belongs to two nonreinforced (ABX and ABY) and two reinforced (AX and AY) configurations. Once again, these differences will result in more positive feedback being directed to the input unit for X than to that for A.

The modification to configural theory can thus explain the results from the two experiments reported above and can also be applied to other discriminations in which a change in

associability has been detected. Dopson, Esber and Pearce. (2010) have shown that the associability of A is greater than that of X after training with an AX+ BX– discrimination, and Dopson, Williams, Esber and Pearce (2010) have shown that this training also results in the associability of B being greater than that of X. Table 4, which is based on exactly the same principles as Table 2, shows how predictions from our modification to configural theory can be derived concerning

Table 4 Calculation of the feedback to Input Units A and X according to the modification to the theory of Pearce (1994) for an AX+ BX– discrimination

		Configural Unit	
		AX+	BX−
Input Unit A		1.1	−0.3
Trial			Feedback per Trial
AX+		1.1	1.1
Total feedback to Input Unit A for entire discrimination:			1.1
Input Unit B			
Trial			Feedback per Trial
BX−			0.3
Total feedback to Input Unit B for entire discrimination:			0.3
Input Unit X			
Trial			Feedback per Trial
AX+		1.1	−0.3
BX−		−1.1	0.3
Total feedback to Input Unit X for entire discrimination:			0.0

the changes in attention to A, B, and X during an AX+ BX– discrimination. Input units A and B are predicted to receive positive feedback, as they are each only connected to configural units that correctly predict the trial outcome whenever these inputs are activated. The absolute magnitude of the associative strength of the AX unit, however, is greater than that of the BX unit, which will result in more positive feedback to A than to B. Table 4 shows also that there will be no net feedback to the input unit for X. Our modification to configural theory thus predicts correctly that more attention will be paid to A and B than to X. The modification also predicts that the associability of A will be greater than that of B.

The new modification to configural theory is speculative and doubtless in need of refinement. Nonetheless, it is difficult to think of an alternative explanation for our results, and the modification can explain the changes in attention to stimuli that occur in relatively simple discriminations. Having said that, one result has already been mentioned that is inconsistent with the modification as it is currently formulated. George and Pearce (1999) trained pigeons with a biconditional discrimination, AB+ CD+ AD– BC–, with a common cue, either X or Y, present on every trial. The proposed modification to configural theory predicts that the same amount of attention will be paid to all of the experimental stimuli, but, as noted earlier, a subsequent test revealed that the birds learned more readily about the previously relevant than about the previously irrelevant stimuli. The implication of this finding, therefore, is that a further revision of the present theoretical proposals will be necessary before they can provide a full account of the role of attention in discrimination learning.

The results from the two experiments we have reported demonstrate that discrimination training may result in animals paying more attention to irrelevant than to relevant stimuli. It is tempting to suggest that this outcome is a consequence of a learning process that normally directs attention to relevant rather than to irrelevant stimuli but that, by virtue of being challenged by an unusual discrimination, was led through its normal operation to direct attention to the wrong stimuli. We have argued that this learning process is based on the formation of configural associations, but the manner in which these associations influence changes in attention remains to be fully understood.

References

- Buhusi, C. V., & Schmajuk, N. A. (1996). Attention, configuration, and hippocampal function. *Hippocampus*, 6, 621–642.
- Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Dissociations of affective and attentional shifting by selective lesions of prefrontal cortex. *Nature*, 380, 69–72.
- Dopson, J. C., Esber, G. R., & Pearce, J. M. (2010). Differences in the 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). Stimuli that signal the absence of reinforcement are paid more attention than are irrelevant stimuli. *Learning & Behavior*, 38, 337–347.
- George, D. N., & Pearce, J. M. (1999). Acquired distinctiveness is controlled by stimulus relevance not correlation with reward. *Journal of Experimental Psychology: Animal Behavior Processes*, 25, 363–373.
- Hall, G. (1991). *Perceptual and associative learning*. Oxford: Oxford University Press, Clarendon Press.
- 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.
- Kirk, R. E. (1968). *Experimental design: procedures for the behavioral sciences*. Belmont: Brooks/Cole.
- Le Pelley, M. E. (2004). The role of associative history in models of associative learning: a selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology*, 57B, 193–243.
- Le Pelley, M. E., & McLaren, I. P. L. (2003). Learned associability and associative change in human causal learning. *The Quarterly Journal of Experimental Psychology*, 56B, 68–79.
- Mackintosh, N. J. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276–298.
- Mackintosh, N. J., & Little, L. (1969). Intradimensional and extradimensional shift learning by pigeons. *Psychonomic Science*, 14, 5–6.
- Moore, J. W., & Stickney, K. J. (1985). Antiassociations: conditioned inhibition in attentional-associative networks. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Conditioned inhibition*. Hillsdale: Erlbaum.
- Pearce, J. M. (1987). A model of stimulus generalization for Pavlovian conditioning. *Psychological Review*, 94, 61–73.
- Pearce, J. M. (1994). Similarity and discrimination: a selective review and a connectionist model. *Psychological Review*, 101, 587–607.
- Pearce, J. M. (2002). Evaluation and development of a connectionist theory of configural learning. *Animal Learning & Behavior*, 30, 73–95.
- Pearce, J. M., Esber, G. R., George, D. N., & Haselgrove, M. (2008). The nature of discrimination learning in pigeons. *Learning & Behavior*, 36, 188–199.
- Pearce, J. M., George, D. N., & Redhead, E. S. (1998). The role of attention in the solution of conditional discriminations. In N. A. Schmajuk & P. C. Holland (Eds.), *Occasion setting: Associative learning and cognition in animals* (pp. 249–275). Washington, DC: American Psychological Association.
- Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not 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 C. J. Mitchell & M. E. Le Pelley (Eds.), *Attention and associative learning: From brain to behaviour* (pp. 11–39). Oxford: Oxford University Press.
- 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.

- Roberts, A. C., Robbins, T. W., & Everitt, B. J. (1988). The effects of intradimensional and extradimensional shifts on visual discrimination learning in humans and non-human primates. *The Quarterly Journal of Experimental Psychology*, 40B, 321–341.
- Schmajuk, N. A., & Moore, J. W. (1985). Real-time attentional models for classical conditioning and the hippocampus. *Physiological Psychology*, 13, 278–290.
- Shepp, B. E., & Eimas, P. D. (1964). Intradimensional and extradimensional shifts in the rat. *Journal of Comparative and Physiological Psychology*, 57, 357–364.
- Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press.
- Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: application of a theory. In R. A. Boakes & M. S. Halliday (Eds.), *Inhibition and learning* (pp. 301–336). New York: Academic Press.
- Williams, D. A., Gawel, J. D., Reimer, D. S., & Mehta, R. (2005). Resistance to interference in complex negative patterning. *Learning & Behavior*, 33, 417–427.
- Williams, D. A., Mehta, R., & Dumont, J. L. (2004). Conditions favoring superconditioning of irrelevant conditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 30, 148–159.
- Williams, D. A., Mehta, R., Poworczyk, T. M., Oriehel, J. S., George, D. N., & Pearce, J. M. (2002). Acquisition of superexcitatory properties by an irrelevant background stimulus. *Journal of Experimental Psychology: Animal Behavior Processes*, 28, 284–297.

This research was supported by a grant from the U.K. Biotechnology and Biological Sciences Research Council. The manuscript was prepared while J.M.P. was a Humboldt Research Fellow at Philipps University, Marburg. He is grateful to Harald Lachnit for the generous provision of support and facilities during this visit.