

Overshadowing and associability change: Examining the contribution of differential stimulus exposure

Peter M. Jones & Mark Haselgrove

Learning & Behavior

ISSN 1543-4494

Volume 41

Number 1

Learn Behav (2013) 41:107-117

DOI 10.3758/s13420-012-0089-z

Learning & Behavior

VOLUME 41, NUMBER 1 ■ MARCH 2013

L&B

EDITOR

Geoffrey Hall, *University of York and University of New South Wales*

ASSOCIATE EDITORS

Jonathon D. Crystal, *Indiana University*

Randolph C. Grace, *University of Canterbury*

Karen L. Hollis, *Mount Holyoke College*

A PSYCHONOMIC SOCIETY PUBLICATION

www.psychonomic.org

ISSN 1543-4494

 Springer



 Springer

Your article is protected by copyright and all rights are held exclusively by Psychonomic Society, Inc.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Overshadowing and associability change: Examining the contribution of differential stimulus exposure

Peter M. Jones · Mark Haselgrove

Published online: 7 September 2012
© Psychonomic Society, Inc. 2012

Abstract In two appetitive conditioning experiments with rats, we investigated the mechanisms responsible for demonstrations of the superior associability of overshadowed conditioned stimuli (CSs) relative to control CSs. In Experiment 1, we investigated whether previous demonstrations were a consequence of differences in the relationship between the CSs and the unconditioned stimulus (US) or of differences in the conditions of exposure to the CSs. Rats received trials with X, Y, and an AB compound, but no delivery of the US (X–, Y–, AB–). A subsequent AY–, AX+, BY + test discrimination revealed that the AY/BY component of the discrimination was solved more readily than the AY/AX component—suggesting the contribution of an exposure effect. In Experiment 2, we better equated the conditions of exposure between A and Y by using AB+, XY+, X– training in Stage 1. In Stage 2, instrumental responses were rewarded during an AY compound. A final test revealed that Y took better control of instrumental responding than did A. The results of these experiments are discussed in terms of classical and contemporary theories of learning and attention.

Keywords Overshadowing · Cue competition · Associability · Associative learning · Attention · Latent inhibition

P. M. Jones
School of Psychology, Cardiff University,
Cardiff, UK

M. Haselgrove (✉)
School of Psychology, University Park, University of Nottingham,
Nottingham NG7 2RD, UK
e-mail: mark.haselgrove@nottingham.ac.uk

A well-documented feature of conditioning is the sensitivity that the conditioned stimulus (CS) shows to the presence of other stimuli during training. This sensitivity, termed *cue competition*, is well exemplified by an effect known as *overshadowing*. In a typical demonstration of overshadowing, two CSs are presented simultaneously and followed on each trial by an unconditioned stimulus (US). In subsequent test trials, the conditioned response to the individual conditioned stimuli is weaker than the response either to the compound of the two (Pavlov, 1927, p. 141) or to a control stimulus that has been conditioned alone (e.g., Kamin, 1969).

One explanation for overshadowing, and indeed of other cue-competition phenomena, appeals to variations in the processing of the CS as a consequence of conditioning, and one of the most successful of these theories is Mackintosh's (1975). According to this theory, a CS will enjoy an enhancement in the extent to which it is processed, or attended to, if it is a better predictor of the US than are all other CSs present on that trial. This notion was formalized by stating that the learning rate parameter for the CS (α) increases if its prediction error¹ is less than the total prediction error for all remaining CSs on that trial. Conversely, a CS will suffer a loss of attention if it is no better a predictor of the US than are the other CSs present on that trial. Again, this notion was formalized: α will decrease if the prediction error of the CS is, at best, equal to the total prediction error of the remaining CSs. Applying Mackintosh's (1975) theory to overshadowing: a CS paired, in isolation, with the US will be the best predictor of the US (relative to, say, contextual cues). α to this CS will therefore increase, permitting the CS to better enter into an association with the US on each trial and to consequently evoke a strong conditioned response. A CS that is conditioned

¹ *Prediction error* is defined as the arithmetic difference between the magnitude of the US (λ) and the expectation of the US (associative strength, V).

in compound with another stimulus, however, will be no better a predictor of the US than is its companion. Consequently, α for this CS will decrease, limiting the extent to which the CS will enter into an association with the US, restricting the CS's ability to evoke a conditioned response.

In a number of experiments conducted during the 1970s (Mackintosh, 1971, 1976; Mackintosh & Reese, 1979), Mackintosh investigated overshadowing and demonstrated some properties that were consistent with his theory of conditioning and attention. More recently, a series of experiments reported by Jones and Haselgrove (2011) tested the predictions of Mackintosh's (1975) analysis of overshadowing. They examined whether a cue that had a history of overshadowing training would subsequently be less successful in taking control of conditioned behavior than was a cue with a history of being conditioned in isolation. In their Experiment 1, rats first received training in which cues X and Y, presented separately, and the compound of cues A and B, presented simultaneously, were followed by a US (X+, Y+, AB+). Following this training, the rats were required to solve a discrimination in which a compound of A and Y was followed by the US (AY+), but compounds of A and X and of B and Y were not (AX-, BY-). According to Mackintosh (1975), the subdiscrimination between AY and BY should have been difficult to solve, as it was based on A and B, which, as a consequence of the prior AB + training, should command little attention. In contrast, the discrimination between AY and AX should have been easier to solve, as it was based on X and Y, which, being the best available predictors of the US during the prior training, should command more attention. The results revealed the opposite pattern: The discrimination between AY and BY was solved more readily than the discrimination between AY and AX, a result that was inconsistent with the idea that attention was *lower* to A and B than to X and Y.

The results of Jones and Haselgrove's (2011) Experiment 2 expanded upon the results of their Experiment 1: Following identical training in Stage 1 (X+, Y+, AB+), rats were presented with an AY compound. Instrumental responses made on one lever (Response 1: R1) during AY led to delivery of the US, but instrumental responses on another lever (Response 2: R2) did not. In a final test, A and Y were presented in isolation, and the number of R1 and R2 responses were measured in extinction. The results showed that rats made more correct responses (R1) during A than they did during Y. This result was again inconsistent with the idea that attention was less to A than to Y; if this were the case, the more salient Y should have taken better control over instrumental responding on R1 than had the less salient A.

Jones and Haselgrove (2011) suggested that their results, while incompatible with Mackintosh's (1975) original model of attention and learning, could be explained by a more complex hybrid model proposed by Pearce and Mackintosh (2010). To avoid undue repetition, the details of this

explanation will not be described fully here, but, in brief, the analysis is based on the idea that an algorithm that employs a *total* prediction error (Rescorla & Wagner, 1972) is used to update associative connections between the CS and the US and that an algorithm that employs an *individual* prediction error between each CS and the US (Bush & Mosteller, 1951) is used to update attention to a CS. Applying a total prediction error to AB + training will limit the associative strength of these two cues to a subasymptotic level. Consequently, the individual prediction error of A and B will be larger than that of X and Y, which, because they were conditioned in isolation, will have asymptotic associative strength. If it is accepted that attention to a CS is determined, at least in part, by the magnitude of its individual prediction error (Pearce & Mackintosh, 2010), then it is possible that attention will be higher to an overshadowed CS than to a control CS. With certain parametric assumptions granted, therefore, Pearce and Mackintosh's theory is able to predict that attention will be higher to overshadowed than to control cues.

However, an unsatisfactory feature of the experiments described by Jones and Haselgrove (2011), and indeed of most studies of overshadowing (but see Dwyer, Haselgrove, & Jones, 2011), was the control stimuli against which the overshadowed stimuli were compared. Unlike the overshadowed stimuli (A and B), the control stimuli (X and Y) were presented in isolation throughout training. Studies of latent inhibition have shown that the associability of CSs is influenced by whether preexposure was conducted in compound with another stimulus or in isolation (e.g., Honey & Hall, 1988, 1989). It therefore remains unclear whether, at test, the differences in performance and discrimination between the stimuli used by Jones and Haselgrove were a consequence of (a) the difference in the relationship that these CSs had with the US, as can be derived from the Pearce–Mackintosh (2010) theory, or (b) whether the CSs were exposed in compound or in isolation. We describe here two experiments through which we attempted to uncouple these two possibilities. To foreshadow the results of these experiments, Experiment 1 demonstrated that mere exposure to an AB compound and to X and Y, in the absence of reinforcement, results in better discrimination between AY and BY than between AY and AX. In Experiment 2, we better equated the conditions of exposure of the overshadowing and control stimuli during Stage 1 by giving rats training of the form AB + XY + X-. Subsequently, instrumental responding on R1, but not R2, was rewarded during an AY compound. Choice tests conducted in extinction between R1 and R2 revealed that Y had taken better control over behavior than had A.

Experiment 1

In their Experiment 1, Jones and Haselgrove (2011) gave rats conditioning trials in which an AB compound and X

and Y were paired with food (AB+, X+, Y+). Following this training (and a successful demonstration that responding was greater to X and Y than to A and B), the animals were required to solve an AY+, AX-, BY- test discrimination. The results of this test revealed that the AY/AX component of the discrimination was solved more slowly than the AY/BY component. One way in which to test whether compound versus individual exposure contributed to the difference in the associability of A/B and X/Y would be to simply replicate Experiment 1 of Jones and Haselgrove, but withhold the delivery of the US following trials with AB, X, and Y during training. Under these circumstances, A and B cannot enter into a different relationship with the US to X and Y, as no USs are delivered during this stage of the experiment. Consequently, in a subsequent AY+, AX-, BY- test discrimination, if the AY/BY component were again to be discriminated more readily than the AY/AX component, it would imply the contribution of compound exposure. Unfortunately, such a direct replication would be unlikely to be sufficiently sensitive to detect differences between the two subdiscriminations. As a consequence of the Stage 1 training, in which rats were given AB- X- Y- training, performance at the outset of the test discrimination would be expected to be low to each compound. AY + training would permit conditioned responding to be acquired to this compound, but the absence of any US on the AX- and BY- trials would ensure that performance to these compounds would remain at the same, low level throughout testing. This situation would be rather different from the conditions of the test discrimination in Jones and Haselgrove's Experiment 1, in which conditioned responding to AX and BY started at a high level (as a consequence of the AB+, X+, Y + training) and gradually reduced during the test stage to a low level—permitting a comparison of AX and BY across the whole scale of performance. To address this issue, we made one small change to the structure of the test discrimination: Throughout testing, the discrimination took the form AY-, AX+, BY+. The logic remained the same; if the associability of A and B was greater than the associability of X and Y, the AY/BY discrimination should be solved more readily than the AY/AX discrimination. However, by switching the compound-US contingencies, we could now examine the acquisition of performance to the crucial AX and BY compounds, again across the whole scale of performance.

Method

Subjects A group of 32 naïve, male Lister hooded rats (*Rattus norvegicus*) served as the subjects. These were supplied by Harlan Olac (Bicester, Oxon., England) and were housed in pairs in a light-proof holding room at the University of Nottingham that was illuminated for 12 h/day. At the time of their arrival, their weight range was 200–225 g. The

rats were subsequently fed a restricted quantity of food daily, so that each rat's weight was reduced to not less than 85 % of its original value. In addition, the minimum permissible weight for each animal was increased gradually in proportion to the weight increases of a separate group of rats, housed in the same room, that were allowed access to food ad libitum. During the experiment, rats received food rations immediately following each experimental session.

Apparatus The experimental sessions took place in eight conditioning chambers (MED Associates, St. Albans, VT; 30.0 × 24.0 × 20.5 cm high) that were housed in separate light- and sound-attenuating cubicles. Ventilation and background noise of 70 dB were provided by an exhaust fan incorporated into the wall of each cubicle. The two smaller walls of each experimental chamber were made of aluminium; the two larger walls (one of which was hinged and served as a door) and the ceiling were made of clear acrylic; and the floor was composed of a series of 19 steel rods running parallel to the smaller walls of the chamber. Each of these rods was 4.8 mm thick, and the distance between them was 1.6 cm, measured center to center. One of the shorter walls was equipped with a recessed magazine into which food could be delivered, and two retractable levers were located to either side of the magazine. The magazine was square shaped, 50 mm wide, and positioned equidistant from the two adjacent walls and 18 mm above the grid floor at its lowest edge. A food hopper was connected to this magazine for the delivery of 45-mg grain-based food pellets (traditional formula; P.J. Noyes, Lancaster, NH). An infrared beam was sent from one lateral side of the magazine and received on the other, and each interruption of this beam could be recorded as magazine activity. The retractable levers (MED Associates), when extended, were 4.8 cm wide and 0.2 cm thick and protruded 1.9 cm from the wall. The horizontal distance from the midpoint of each lever to the midpoint of the wall on which they were mounted was 7.9 cm, and each lever was 6.2 cm from the grid floor. Each lever could either be presented to the rat or fully retracted into the wall, and this state was controlled by the experimental computer. Both levers remained retracted throughout Experiment 1. The opposite wall incorporated three loudspeakers: two square-shaped loudspeakers of 70-mm width that were located in the upper corners of the wall, and a rectangular loudspeaker measuring 35 × 70 mm that was located 35 mm from the ceiling and equidistant from the two adjacent walls. Each of these loudspeakers was used for the delivery of auditory stimuli. 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 period of silence, and could also emit a 74-dB “buzzer” composed of a 500-Hz train of clicks. The upper right loudspeaker could emit a 78-dB white noise, and the rectangular loudspeaker could emit a 77-dB,

2.9-kHz tone. Additionally, a relay was attached to the outside of this wall that could be used to produce a 76-dB series of clicks, such that a click occurred every 0.08 s. A total of five auditory stimuli could therefore be presented to the animals: a pulsed 8-kHz tone, a 2.9-kHz tone, white noise, a clicker, and a buzzer. Each of these stimuli was 15 s in duration during Experiment 1. Where two stimuli were presented in compound, they had simultaneous onsets and offsets. Control of all experimental events and recording of magazine activity were conducted by a computer that was programmed with the MED-PC programming language.

Procedure All rats initially received one session of magazine training, which lasted for 30 min. During this session, a food pellet was delivered to the magazine every 60 s. All of the rats consumed all pellets during this session. Before the beginning of the following stage of the experiment, auditory stimuli were assigned to serve as A, B, X, and Y. For half of the animals, the white noise and the 2.9-kHz tone served as A and B, and the clicker and the pulsed 8-kHz tone served as X and Y. For the remaining rats, this arrangement was reversed. Within each of these subgroups, the stimuli assigned to A and B were counterbalanced, as were the stimuli assigned to X and Y. The buzzer served as P (see below) for all animals.

During each of the following 12 sessions, rats received training of the form AB–, X–, Y–. In addition, trials with P + were included to discourage the animal from sleeping throughout the duration of the session. Each session was 52 min long and contained six trials of each type, block-randomized so that each block of eight trials contained two of each trial type. The intertrial interval (ITI), which was defined as the period between the end of one trial and the beginning of the next, varied randomly around a mean for each session of 105 s, with a range of 65–145 s. Presentation of P was immediately followed by the delivery of two food pellets to the magazine; no food pellets were delivered following the AB compound, X, or Y. Following these 12 sessions, rats were trained for two sessions with an AY–, AX+, BY + discrimination. Each of these sessions was 68 min in duration, and each contained 16 presentations of AY and eight presentations of each of AX and BY. Two food pellets were delivered to the magazine following AX and BY, and no food was delivered following AY. Each successive block of eight trials contained four trials with AY and two with each of the other two trial types, in a random order. All 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 experiment. Conditioning with stimulus P proceeded smoothly, and by the final session of

Stage 1, the mean number of responses per minute during P was 32.8. Responding to the average of X and Y (X/Y) and to the AB compound remained at similar, low levels throughout Stage 1, and by the final session of Stage 1, the mean numbers of responses during X/Y and AB were, respectively, 2.1 and 1.9. A one-way analysis of variance (ANOVA) of individual mean responses per minute with stimulus (P, X/Y, AB) as the variable revealed a significant difference among P, X/Y, and AB, $F(2, 62) = 149.80$, $MSE = 67.27$. Post-hoc tests conducted according to the Bonferroni procedure (corrected $p = .017$) revealed differences between P and X/Y and between P and the AB compound, $t(31) > 12.14$, but no difference between X/Y and AB, $t(31) = 0.63$. We also computed, for each animal separately, the greater of the number of responses per minute to X or to Y. The mean of this value was 3.15. The difference between this value and the mean number of responses per minute during the AB compound was not significant, $t(31) = 1.60$. The mean rate of responding during the 15-s intervals prior to the stimuli was 1.3.

The mean numbers of responses per minute during the 15-s intervals prior to the presentation of AX, AY, and BY trials were 4.4 and 4.1 for Sessions 1 and 2, respectively. Figure 1 shows the mean rate of responding on each of the trials with AX and BY and the mean responding on each pair of two successive trials with AY during the two sessions of the test discrimination. The discrimination was successfully solved, as evidenced by the higher rate of responding during the reinforced AX and BY trials than during the nonreinforced AY trials. Interestingly, however, the subdiscrimination between

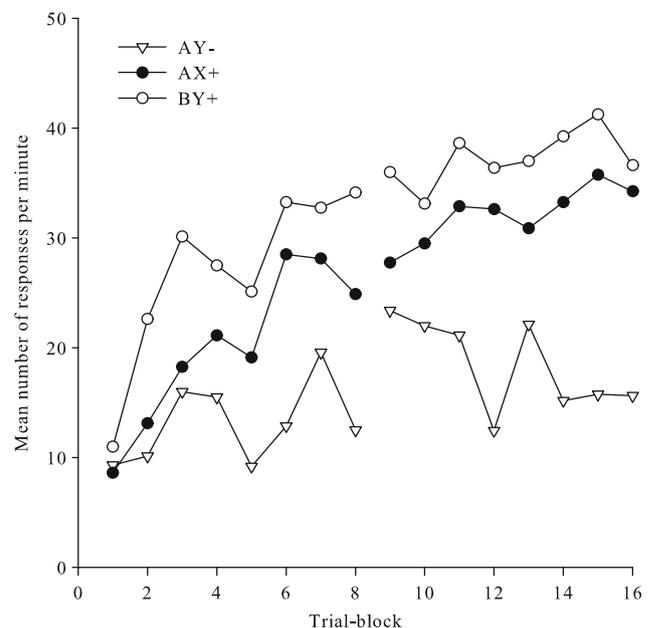


Fig. 1 Results of the test discrimination from Experiment 1: Mean numbers of magazine responses per minute during AY, AX, and BY, on each test trial with AX and BY and on successive two-trial blocks with AY. The + and – symbols refer to food and no food, respectively

AY and BY was solved more readily than the subdiscrimination between AY and AX, as the mean number of responses per minute was higher during BY than during AX. This impression was confirmed with a two-way ANOVA of individual responses per minute with the variables stimulus (AY, AX, BY) and trial (1–16), which revealed a main effect of stimulus, $F(2, 60) = 27.82$, $MSE = 1,263.48$, a main effect of trial, $F(15, 450) = 18.87$, $MSE = 192.94$, and an interaction between these variables, $F(30, 900) = 2.96$, $MSE = 161.44$. Post-hoc tests, again conducted according to the Bonferroni procedure, revealed that the number of responses per minute, averaged across the two sessions, was significantly higher for BY (32.2) than for AX (26.2), which itself was significantly higher than the rate for AY (15.8), $t(31) > 2.30$.

Following nonreinforced presentations of X, Y, and a compound of AB, the discrimination of BY from AY was superior to the discrimination of AX from AY. This pattern of test discrimination is identical to that reported in Jones and Haselgrove (2011), and therefore supports the possibility that the compound exposure to A and B, as compared to the elemental exposure to X and Y, may have contributed to their results. The results of Experiment 1 are also consistent with the results of a latent inhibition experiment reported by Honey and Hall (1988). Using a conditioned flavor aversion procedure, they preexposed one group of rats to flavor A and another group of rats to a compound of flavors A and B. To examine the associability of A, this flavor was subsequently presented to the animal and an aversion conditioned to it. The results showed that conditioning to A proceeded more readily in the animals that had been preexposed to the AB compound than in the animals preexposed to A alone (see also Honey & Hall, 1989, Experiment 1a & b; Lubow, Wagner, & Weiner, 1982, Experiment 2; Mackintosh, 1973; Rudy, Krauter, & Gaffuri, 1976, Experiment 2). The similarity of these results from studies of latent inhibition to the present data is clear: The associability of a CS, whether measured with an acquisition test or a discrimination test, is higher when it has received compound preexposure rather than preexposure in isolation.

It must be noted, however, that other experiments have failed to find any difference in conditioning following either compound preexposure or preexposure in isolation (e.g., Baker & Mercier, 1983; Honey & Hall, 1989, Experiment 2; Mercier & Baker, 1985; Rudy et al., 1976, Experiment 1), while more recent experiments have revealed the opposite result: that conditioning is *less* successful with a stimulus following compound preexposure than following preexposure in isolation (e.g., Hall & Rodriguez, 2011; Leung, Killcross, & Westbrook, 2011; Rodriguez & Hall, 2008). The circumstances that either promote or prevent the attenuation of latent inhibition with compound preexposure have yet to be fully determined. However, Rodriguez and Hall (2008; see also Honey & Hall, 1989) have suggested that preexposing the CS in compound with

another stimulus may cause perceptual interactions that result in the CS being encoded differently during preexposure to conditioning. The effects of preexposure on conditioning will then be attenuated as a consequence of generalization decrement—a possibility that is particularly likely when compound preexposure is conducted with stimuli that are drawn from the same sensory modality (e.g., Honey & Hall, 1989). This analysis is particularly pertinent to the results of the present experiment, as well as to the results of Experiment 1 of Jones and Haselgrove (2011), as both of these experiments examined the associability of auditory CSs following exposure to the CS either in isolation or in compound with another CS. Consequently, it is conceivable that the results of these experiments were a consequence of the combination of the effects of latent inhibition and generalization decrement.

Experiment 2

Experiment 1 revealed that exposure to an AB compound and to X and Y in isolation was sufficient to result in differences in the associability of these stimuli. Following this exposure, during an AY–, AX+, BY + discrimination, responding was significantly higher to the BY compound than to the AX compound. These results suggest that an explanation for the results of the experiments reported by Jones and Haselgrove (2011), which emphasizes the contribution of differences in the effect of compound versus elemental exposure on variations in CS associability, would not be completely without merit. Before this explanation is fully accepted, however, we must consider the possibility that the effects of compound exposure did not contribute *entirely* to the results reported by Jones and Haselgrove. It is possible that a second mechanism contributed to these results, in addition to the effects of compound exposure. It is possible, for example, that differences in the relationships that A and B had with the US relative to X and Y also contributed to the results, in addition to the effect of compound exposure. To investigate this matter, we sought to equate, as best we could, the conditions of exposure of the overshadowed and control stimuli in an attempt to nullify the differential effect of this variable, while at the same time preserving an overshadowing effect. To achieve this, we employed a control for overshadowing that we have used elsewhere (Dwyer et al., 2011). In this design, as before, an AB compound was paired with the US, but rather than conditioning X and Y in isolation, trials were given in which X and Y were presented in compound and paired with the US, while X was presented in isolation: AB+, XY+, X–. A comparison can then be made between A and Y. Presentation of X in the absence of the US should protect Y from overshadowing (e.g., Rescorla & Wagner, 1972), enabling Y to serve as a control stimulus for A. Crucially, however, Y

will have been exposed in compound in the same way as A. Consequently, any differences in associability that may be a result of the conditions of exposure will, to some extent, be nullified.

One possible way to examine differences in attention to A and Y following this training would be to use the same type of compound test discrimination employed in both Jones and Haselgrove's (2011) Experiment 1 and the present Experiment 1. However, this test procedure would require an expansion of the training just described to AB + XY + ZX + X⁻, in order to provide the appropriate number of CSs (A, B, Y, and Z) for the test discrimination. The complete counterbalancing of five cues was not practical with the apparatus used in these studies, and consequently, we retained the simpler AB⁺, XY⁺, X⁻ design used by Dwyer et al. (2011), and employed the same variation on an optional shift test employed by Jones and Haselgrove in their Experiment 2 in order to assess attention to just A and Y. Thus, following Pavlovian conditioning, trials were given in which instrumental responses on R1 (but not R2) were followed by food in the presence of an AY compound. If attention were higher to Y than to A, then Y should take control of performance to a greater extent than A. Thus, on a subsequent choice test between R1 and R2, animals should make more responses on R1 than on R2 during Y, but this discrimination should be less evident during A. If attention were higher to A than to Y, this pattern of responding should be reversed.

Method

Subjects and apparatus The subjects were 32 naïve, male hooded Lister rats that were maintained in the same manner as for Experiment 1. The apparatus was the same as that used in Experiment 1, except that the retractable levers were presented to the animals during some trials, as detailed below.

Procedure The rats initially received a single session of magazine training, according to the same procedure described for Experiment 1. Animals were then trained to press the two levers during four instrumental training sessions. Each of these sessions was 1 h in length and consisted of 20 presentations of each lever, in an alternating fashion. Each presentation was 1 min in length, with a gap of 30 s between each lever presentation. While the levers were available, rats were able to earn food pellets at a rate of reinforcement that was reduced gradually from continuous reinforcement, during the first session, to a variable-interval schedule with a mean interval of 15 s, during the last. They were then trained for 12 sessions with both levers retracted, with an AB⁺, XY⁺, X⁻ discrimination. In keeping with the design of Experiment 2 reported by Jones and Haselgrove

(2011), P⁻ trials were also included in this stage. Other details of these sessions were the same as for the 12 training sessions given during Experiment 1. Following this stage, rats were trained for eight sessions with an instrumental discrimination. Both levers were presented simultaneously during AY and P, and rats were able to earn food by pressing one lever only—a “correct” response. A correct response was defined as a response on R1 during AY and on R2 during P; other responses were considered “incorrect.” For half of the animals, the left lever was assigned as R1 and the right lever as R2, with this arrangement reversed for the remaining rats. These assignments were orthogonal to the assignments of the stimuli, and the two levers were retracted at the end of each trial. Sessions during this stage were 71 min in length and contained 16 trials of each type. Each trial was 20 s in duration, and no food pellets could be earned during the initial 15 s of each trial; the first correct response after this 15-s period (but before the termination of the trial, 5 s later) resulted in the delivery of two food pellets to the magazine. The numbers of responses on R1 and R2 were recorded by the computer during the first 15 s of each trial. Because no pellets were delivered to the magazine during this time, this measure of responding was not compromised by the delivery of the US.

The levels of instrumental control taken by A and Y were then assessed in a final test session. This session was 71 min in length and contained a total of 32 trials. The first eight trials were training trials of the type described above: four trials with AY and four with P, in a random order. Food was delivered following correct responses in the manner described above. The remaining 24 trials were test trials with A and Y. Each stimulus was presented on half of the trials, with both levers available but no food delivered. These trials were block-randomized, so that each block of four trials consisted of two trials of each type. Other details of this test session were the same as for the previous training stage.

Results and discussion

Training in Stage 1 again proceeded smoothly, and by the final session of conditioning, the mean numbers of responses per minute made during P, X, AB, and XY were, respectively, 5.5, 12.8, 40.0, and 39.3. A one-way ANOVA revealed a difference among these means, $F(3, 93) = 60.83$, $MSE = 167.61$. Bonferroni-corrected t tests (corrected p value = .008) revealed that the reinforced compounds (AB and XY) did not differ from each other, $t(31) = 0.34$, but all remaining pair-wise comparisons were significant, $t_s(31) > 3.80$. The mean number of responses made during the 15-s interval prior to these trials was 3.6. By the final session of instrumental conditioning, the mean percentages of responses that were correct during the first 15 s of AY and P were both 72.9 %.

Both of these means were significantly different from chance (50 %), $t(31) > 10.34$, but they did not differ from each other, $t(31) = 0.001$.

Figure 2 shows the mean percentages of responses that were correct during the first 15 s of A and Y across the 6 two-trial blocks of the final test session. From the outset of the test, it is clear that the percentage of correct responses was higher during Y than during A, and throughout the test there was no indication that responding during A was above chance (50 %). This implies that Y took almost complete control of responding. By the end of the test session, presumably because of the repeated nonreinforcement of these two responses, the mean percentage of correct responses during Y approached chance. These impressions were confirmed with a two-way ANOVA of individual percent-correct responses with the variables CS (A vs. Y) and trial block (1–6), which revealed an effect of CS, $F(1, 31) = 21.40$, $MSE = 0.14$, an effect of trial block, $F(5, 155) = 3.21$, $MSE = 0.05$, and an interaction between these variables, $F(5, 155) = 3.33$, $MSE = 0.04$. Simple-effects analysis revealed an effect of trial block for Y, $F(5, 310) = 5.26$, $MSE = 0.04$, but not for A, $F(5, 310) = 1.27$, $MSE = 0.04$, and that A and Y differed from each other on Trial Blocks 1 to 5, $F(1, 186) > 5.61$, $MSE = 0.05$. In addition, the mean percentage of correct responses, pooled across the entire test, was significantly different from chance for Y (63.5 %), $t(31) = 5.35$, but not for A (46.2 %), $t(31) = 1.34$. The number of responses made in the magazine were also recorded during the test session. This measure may be taken to be a Pavlovian conditioned response that allows us to

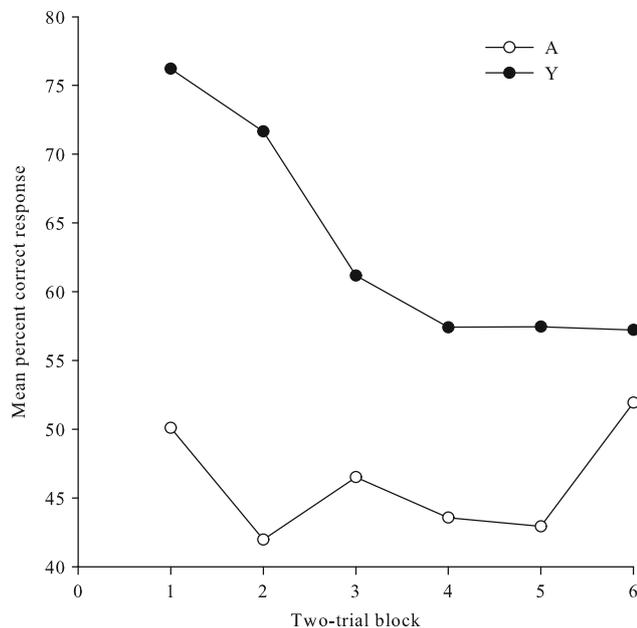


Fig. 2 Results of the final choice test from Experiment 2: Mean percentages of correct leverpress responses made during A and Y across the 6 two-trial blocks of the test session

assess the associative strength acquired by the CSs in Stage 1 of the experiment, and thus to determine whether overshadowing itself was present. If it was, we would expect responding to be higher during Y than for A. Overshadowing was present; the mean numbers of responses per minute made during the first 15 s of the CSs were 13.0 for Y and 9.2 for A. The difference between these means was significant, $t(31) = 6.14$.

The results of this experiment contrast starkly with the results of Experiment 2 reported by Jones and Haselgrove (2011). Using the same apparatus, species, and stimuli, and with training and testing procedures extremely similar to those of the present experiment, Jones and Haselgrove gave rats AB + X + Y + P– training, followed by training in which instrumental responses on R1 (but not R2) were reinforced during an AY compound. Their test results showed that A took better control of instrumental responding than did Y during training in which R1 was reinforced during an AY compound. The present results reveal the opposite pattern of data. Taken together, the present Experiment 2 and Experiment 2 of Jones and Haselgrove imply that associative strength alone does not determine which component of a compound CS takes control of instrumental conditioning. In both of these experiments, standard associative models (e.g., Rescorla & Wagner, 1972) predict that the associative strength of A would be less than that of Y. Instead, it seems likely to us that the crucial difference between these two experiments lies in the way in which the control cue, Y, was treated in Stage 1 of the present experiment. For the present experiment, Y was experienced in compound with another cue (X) in an attempt to better equate the conditions of exposure of this CS with those of A. Under these circumstances, the associability of an overshadowed CS was less than that of a control CS.

General discussion

In two appetitive conditioning experiments with rats, we investigated why previous experiments conducted in our laboratory (Jones & Haselgrove, 2011) had demonstrated that attention is seemingly superior to CSs with a history of overshadowing training, relative to CSs with a history of conditioning in isolation. The results of Jones and Haselgrove's experiments are at variance with the predictions of classical (e.g., Mackintosh, 1975; Pearce & Hall, 1980) and more contemporary (e.g., Esber & Haselgrove, 2011; Le Pelley, 2004) analyses of the role of predictiveness in attention. As was noted in the introduction, Mackintosh's theory predicts that attention will increase to CSs that are the best available predictor of the US, but that attention will decrease to CSs that are, at best, as good a predictor of the US as are other available CSs. It follows from this analysis, therefore, that attention will be lower to overshadowed cues than to cues conditioned in

isolation.² According to Esber and Haselgrove's theory, the salience of a CS (and therefore the attention that it can attract) is, in part, related to its association with motivationally significant outcomes (such as US, and no-US, representations; Konorski, 1967). A CS conditioned in isolation will have higher associative strength with the US than will a cue conditioned in compound, and it therefore follows from this theory that attention will be lower to an overshadowed CS than to a CS conditioned in isolation. Le Pelley's theory proposes that the attention paid to a CS is the product of predictiveness (Mackintosh, 1975) and uncertainty (Pearce & Hall, 1980) mechanisms. As neither the Mackintosh nor the Pearce–Hall theories, in isolation, can formulate an explanation for the results described by Jones and Haselgrove, Le Pelley's theory similarly struggles. Although the hybrid model proposed by Pearce and Mackintosh (2010) can, in principle, provide an explanation for the data described by Jones and Haselgrove, it relies on parametric assumptions that have yet to be substantiated. Therefore, how appropriate this analysis of Jones and Haselgrove's results is remains to be determined.

The present experiments go some way toward relieving the tension that exists between the predictions of a variety of theories and the empirical observations of Jones and Haselgrove (2011). The present Experiment 1 showed that, even in the absence of the delivery of explicit USs, mere exposure to an AB compound and to X and Y in isolation resulted in A and B being more discriminable than X and Y. This result is consistent with the notion that A and B possess higher attention than do X and Y, and also with prior studies of latent inhibition (e.g., Honey & Hall, 1988, 1989), which have shown faster conditioning to A following nonreinforced pre-exposure to AB than to A alone. It seems, therefore, that a mechanism sensitive to the different conditions of exposure that A and B receive relative to X and Y would constitute an appropriate analysis for the results of our earlier experiments. This notion is supported by the results of Experiment 2, which showed that when the conditions of exposure were better matched during Pavlovian conditioning, an overshadowed CS was less able to serve as a discriminative stimulus for instrumental conditioning than was a control CS—a result in keeping with the idea that an overshadowed cue captures *less* attention than does a cue that is a good predictor of the US.

It remains, then, to discuss the nature of the mechanism that permits greater associability to a CS that has a history of exposure in compound with another stimulus, relative to a CS that does not. In discussing their latent inhibition results, Honey and Hall (1988) considered two possible

explanations for why conditioning with A is less attenuated following exposure to AB than to A: One explanation was perceptual, the other associative. The perceptual explanation proposed that whatever mechanism produces a decrement in attention as a consequence of exposure (and therefore generates latent inhibition) will only be effective if the CS presented during preexposure is identified by the organism as being the same as that presented during conditioning. If it is not, then a generalization decrement will occur, and latent inhibition will be incomplete (Siegel, 1969). Honey and Hall (1989) suggested that this generalization decrement occurs because A is perceived differently when in compound with B than when it is presented alone (see also Wagner, 2003; Wagner & Brandon, 2001). Honey and Hall (1989) supported this analysis by demonstrating that when A and B were drawn from different sensory modalities (and the perception of A was less likely to be influenced by the presence of B), conditioning to A proceeded at the same rate, irrespective of whether preexposure consisted of trials with an AB compound or with A alone. This explanation seems relevant to the present Experiment 1 and to the experiments described by Jones and Haselgrove (2011), all of which employed stimuli from the same, auditory, modality. Thus, during the AB+, X+, Y + conditioning trials, latent inhibition may have accrued in equal measure to A, B, X, and Y (Hall & Pearce, 1979), but during subsequent tests in which the associability of these CSs was assessed, presenting A in the absence of B (and vice versa) restored attention to these CSs, as they were, to a degree, identified as stimuli different from those experienced during conditioning. As far as the present Experiment 2 is concerned, however, conditioning in Stage 1 ensured that A was exposed with B in a manner comparable to the way in which Y was exposed with X. Consequently, the generalization decrements encountered by A and Y were similar when they were subsequently experienced as a compound in Stage 2, and the effects of latent inhibition were equalized. The obvious prediction from this analysis is that conditioning with AB+X+Y+ trials using cues drawn from different sensory modalities should produce a result much more in keeping with the present Experiment 2 than with the Experiment 2 reported by Jones and Haselgrove (2011). This experiment remains to be conducted.

The preceding discussion, however, does give rise to a further puzzle. The present Experiment 2 provided evidence consistent with the idea that Y acquired more attention than did A by virtue of its being a better predictor of the US (e.g., Esber & Haselgrove, 2011; Le Pelley, 2004; Mackintosh, 1975). Presumably, the same effect was also present in the experiments described by Jones and Haselgrove (2011). This being the case, it suggests that although the mechanism that (through exposure) reduces attention to a CS is sensitive to the effect of generalization decrement, the mechanism

² The Pearce and Hall (1980) model predicts that conditioning with the AB compound will proceed more rapidly than conditioning with X or Y in isolation. Prior to asymptote being reached, therefore, the total prediction error (and, therefore, α) of A and B will be smaller than that of X or Y. The theory proposed by Pearce and Hall therefore makes the same prediction as Mackintosh's (1975) theory.

that *increases* attention to CSs by virtue of their being good predictors of the US is not. If it were, then in Jones and Haselgrove's previous experiments, the attention acquired by X and Y as a consequence of their status as good predictors of the US should have transferred more successfully to their subsequent tests than did the attention acquired by A and B—counteracting the effect of the differential transfer of latent inhibition. It is possible, then, that the acquisition of attention by a CS as a consequence of establishing it as a good predictor of the US is particularly tolerant to the effects of generalization decrement. In a sense, this tolerance is already known of from studies of the intradimensional/extradimensional shift effect (e.g., George & Pearce, 1999; Mackintosh & Little, 1969), where the difference in attention paid to the relevant and irrelevant stimuli from AX + BX– training is seen to generalize well to other stimuli from the same dimension. The reason why increases and decreases in attention may be differentially tolerant to generalization decrement remains, for the time being, undetermined.

The associative explanation proposed by Honey and Hall (1988) to account for their latent inhibition results draws on the theory of learning and memory proposed by Wagner (1981). According to this analysis, excitatory associative learning takes place between stimuli when they are concurrently in a central state of activation (A1). Repeated non-reinforced preexposure of a CS will permit the establishment of an association between elements of the context and the CS. According to Wagner (1981), this arrangement will enable the context to activate elements of the CS into a more peripheral state of activation (A2) that does not permit excitatory associative learning. Subsequent conditioning trials with the CS will therefore proceed slowly, due to elements of the CS being in a state that does not support conditioning. According to Wagner (1981), the number of elements that can be in the A1 state at any moment is limited. As a consequence of this, exposure to a compound of CSs during preexposure will permit only a proportion of each CS's elements to enter the A1 state, and thus to be associated with the context. Thus, upon subsequent conditioning in isolation, some of the elements of the CS will not be primed into the A2 state by the context, and will therefore be amenable to enter the A1 state, and hence be associated with the US. This analysis can also be applied to the experiments described by Jones and Haselgrove (2011). X+, Y + trials will ensure that the context will enter well into associations with X and with Y, thus priming elements of these stimuli into the A2 state, retarding their ability to subsequently enter into associative relations in the test discrimination. However, the concurrent presentation of two stimuli on AB + trials will restrict the extent to which the context can activate A and B into the A2 state—affording these stimuli processing in the A1 state, and thus giving them the freedom to subsequently enter into associative relations.

A slight shortcoming of the preceding analysis, however, is the conflict it creates with the observation of overshadowing itself. According to Wagner's (1981) theory, overshadowing is also a consequence of the restriction on the number of elements that can, at any one time, be activated in the A1 state. When only one CS is paired with the US (e.g., X+), elements of both of these stimuli may concurrently be in the A1 state—permitting their association. However, if a compound of two CSs is paired with the US (e.g., AB+), then not all of the elements of the two CSs (or, indeed, the US) will concurrently be in the A1 state. The consequence of this will be a weaker association with either of the CSs and the US, and thus overshadowing. However, if the context is better able to prime elements of X into the A2 state than either A or B, this process will be undermined, for now more elements of A and B than of X will be able to enter the A1 state during conditioning. Presumably, variations in the parameters of the model will permit the effect of capacity limitation on the CS–US association to dominate over the effect of capacity limitation on the context–CS association, and thus reveal overshadowing, while still permitting the model to predict that the associability of A and B will be more substantial than that of X and Y. What the range of these parameters is and whether it can be justified remains to be seen. Furthermore, to the best of our knowledge, the theory suggested by Wagner (1981) provides no basis for expecting that an overshadowed CS will be less able to take control over instrumental conditioning than a control CS. It seems to lack, therefore, an explanation for the results of Experiment 2.

A third possible explanation for the differential associability of CSs that have a history of exposure in isolation and in compound draws its inspiration from an effect first reported by Hall and Pearce (1982). In a series of experiments, Hall and Pearce demonstrated that conditioning of a CS–strong shock association was attenuated if this training was preceded by CS–weak shock training. However, this effect was attenuated if extinction trials were given between conditioning with the weak and strong shocks. The analysis provided by Hall and Pearce (1982) for these effects followed from the Pearce and Hall (1980) model: Conditioning with the weak shock resulted in a reduction in the associability of the CS, retarding subsequent conditioning with the strong shock, but the surprising omission of the US during extinction restored associability. It is possible that in the experiments reported by Jones and Haselgrove (2011) and in the present Experiment 1 that associations formed not only between the CSs and the US, but also between simultaneously presented CSs (Speers, Gillan, & Rescorla, 1980). This being the case, presenting A in the absence of B (and vice versa) during the subsequent tests for associability may have constituted a surprising event that restored the associability of these CSs. It is difficult to know how seriously to

take this analysis, not least because the original formulation of the Pearce–Hall theory was concerned with variations in CS associability that were a consequence of predictive accuracy—and thus serial associations, not simultaneous associations. However, it is conceivable that a model that emphasizes the role of “predictive” accuracy in both simultaneous and sequential associative learning might be successful (e.g., Schmajuk, Lam, & Gray, 1996).

Whatever the virtue of these various mechanisms, both the present experiments and the experiments reported by Jones and Haselgrove (2011) suggest that the influence of overshadowing training on CS associability is multiply determined. On the one hand, it seems that by having a poor predictive relationship with the US, overshadowed CSs have a lower associability than do control CSs. However, on the other hand, unless carefully controlled for, differences in the conditions of exposure can reverse this effect.

Author note This research was supported by BBSRC New Investigator Grant No. BB/F01239X/1 to M.H.

References

- Baker AG, Mercier P (1983) Prior experience with the conditioning events: Evidence for a rich cognitive representation. In: Wagner AR, Herrnstein R, Commons M (eds) *Quantitative analysis of behavior: Acquisition processes*, vol 3. Ballinger, Cambridge, pp 117–143
- Bush RR, Mosteller F (1951) A mathematical model for simple learning. *Psychological Review* 58:313–323
- Dwyer DM, Haselgrove M, Jones PM (2011) Cue interactions in flavor preference learning: A configural analysis. *Journal of Experimental Psychology: Animal Behavior Processes* 37:41–57
- Esber GR, Haselgrove M (2011) Reconciling the influence of predictiveness and uncertainty on stimulus salience: A model of attention in associative learning. *Proceedings of the Royal Society B* 278:2553–2561
- George DN, Pearce JM (1999) Acquired distinctiveness is controlled by stimulus relevance not correlation with reward. *Journal of Experimental Psychology: Animal Behavior Processes* 25:363–373
- Hall G, Pearce JM (1979) Latent inhibition of a CS during CS–US pairings. *Journal of Experimental Psychology: Animal Behavior Processes* 5:31–42. doi:10.1037/0097-7403.5.1.31
- Hall G, Pearce JM (1982) Restoring the associability of a pre-exposed CS by a surprising event. *Quarterly Journal of Experimental Psychology* 34B:127–140
- Hall G, Rodriguez G (2011) Blocking of potentiation of latent inhibition. *Journal of Experimental Psychology: Animal Behavior Processes* 37:127–131
- Honey RC, Hall G (1988) Overshadowing and blocking procedures in latent inhibition. *Quarterly Journal of Experimental Psychology* 40B:163–186
- Honey RC, Hall G (1989) Attenuation of latent inhibition after compound pre-exposure: Associative and perceptual explanations. *Quarterly Journal of Experimental Psychology* 41B:355–368
- Jones PM, Haselgrove M (2011) Overshadowing and associability change. *Journal of Experimental Psychology: Animal Behavior Processes* 37:287–299
- Kamin LJ (1969) Predictability, surprise, attention, and conditioning. In: Campbell BA, Church RM (eds) *Punishment and aversive behavior*. Appleton-Century-Crofts, New York, pp 279–296
- Konorski J (1967) *Integrative activity of the brain*. University of Chicago Press, Chicago
- Le Pelley ME (2004) The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology* 57B:193–243. doi:10.1080/02724990344000141
- Leung HT, Killcross AS, Westbrook F (2011) Additional exposures to a compound of two preexposed stimuli deepen latent inhibition. *Journal of Experimental Psychology: Animal Behavior Processes* 37:394–406
- Lubow RE, Wagner M, Weiner I (1982) The effects of compound stimulus preexposure of two elements differing in salience on the acquisition of conditioned suppression. *Animal Learning and Behaviour* 10:485–489
- Mackintosh NJ (1971) An analysis of overshadowing and blocking. *Quarterly Journal of Experimental Psychology* 23:118–125
- Mackintosh NJ (1973) Stimulus selection: learning to ignore stimuli that predict no change in reinforcement. In: Hinde RA, Hinde JS (eds) *Constraints on learning*. Academic Press, London, pp 75–96
- Mackintosh NJ (1975) A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review* 82:276–298. doi:10.1037/h0076778
- Mackintosh NJ (1976) Overshadowing and stimulus intensity. *Animal Learning and Behavior* 4:186–192
- Mackintosh NJ, Little L (1969) Intradimensional and extradimensional shift learning by pigeons. *Psychonomic Science* 14:5–6
- Mackintosh NJ, Reese B (1979) One-trial overshadowing. *Quarterly Journal of Experimental Psychology* 31:519–526
- Mercier P, Baker AG (1985) Latent inhibition, habituation, and sensory preconditioning: A test of priming in short-term memory. *Journal of Experimental Psychology: Animal Behavior Processes* 11:485–501
- Pavlov IP (1927) *Conditioned reflexes* (G. V. Anrep, Trans.). Oxford University Press, London
- Pearce JM, Hall G (1980) A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review* 87:532–552. doi:10.1037/0033-295X.87.6.532
- Pearce JM, Mackintosh NJ (2010) Two theories of attention: A review and a possible integration. In: Mitchell CJ, Le Pelley ME (eds) *Attention and associative learning: From brain to behaviour*. Oxford University Press, Oxford, pp 11–39
- Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF (eds) *Classical conditioning II: Current research and theory*. Appleton-Century-Crofts, New York, pp 64–99
- Rodriguez G, Hall G (2008) Potentiation of latent inhibition. *Journal of Experimental Psychology: Animal Behavior Processes* 34:352–360
- Rudy JW, Krauter EE, Gaffuri A (1976) Attenuation of the latent inhibition effect by prior exposure to another stimulus. *Journal of Experimental Psychology: Animal Behavior Processes* 2:235–247
- Schmajuk NA, Lam Y-W, Gray JA (1996) Latent inhibition: A neural network approach. *Journal of Experimental Psychology: Animal Behavior Processes* 22:321–349. doi:10.1037/0097-7403.22.3.321

- Siegel S (1969) Generalization of latent inhibition. *Journal of Comparative and Physiological Psychology* 69:157–159
- Speers MA, Gillan DJ, Rescorla RA (1980) Within-compound associations in a variety of compound conditioning procedures. *Learning and Motivation* 11:135–149
- Wagner AR (1981) SOP: A model of automatic memory processing in animal behavior. In: Spear NE, Miller RR (eds) *Information processing in animals: Memory mechanisms*. Erlbaum, Hillsdale, pp 5–47
- Wagner AR (2003) Context-sensitive elemental theory. *Quarterly Journal of Experimental Psychology* 56B:7–29
- Wagner AR, Brandon SE (2001) A componential theory of Pavlovian conditioning. In: Mower RR, Klein SB (eds) *Handbook of contemporary learning theories*. Erlbaum, Mahwah, pp 23–64