Modulation of attention in discrimination learning: The roles of stimulus relevance and stimulus–outcome correlation

Metin Uengoer · Harald Lachnit

Published online: 2 October 2011
© Psychonomic Society, Inc. 2011

Abstract The advantage in learning after an intradimensional shift rather than an extradimensional shift has been widely used as a behavioural marker of attentional changes during discrimination learning in different fields of neuroscientific study. However, some of the factors assumed to guide these attentional changes have not been completely disentangled by previous research. In two predictive-learning experiments, we investigated the importance of stimulus relevance and of stimulus–outcome correlation for the modulation of attention. In each experiment, participants were trained on two discrimination problems given in successive order. Each problem required participants to differentiate stimuli varying on two dimensions. We found that acquisition of the second discrimination was influenced by whether its relevant dimension (*Exp. 1*) or its irrelevant dimension (*Exp. 2*) had previously been trained as relevant and uncorrelated or as irrelevant and uncorrelated. We also observed that acquisition of the second discrimination was independent of whether its relevant dimension (*Exp. 1*) or its irrelevant dimension (*Exp. 2*) had previously been trained as relevant and uncorrelated or as relevant and correlated. Our results indicated that the modulation of attention is guided by stimulus relevance and not by stimulus–outcome correlation.

Keywords Associative learning · Attention · Discrimination · Human learning · Intradimensional · Extradimensional

Discrimination learning is thought to involve changes in the amount of attention paid to the stimuli present during training (for brief surveys, see Hall, 1991; Le Pelley, 2004; Pearce & Mackintosh, 2010). One line of evidence for the modulation of attention in discrimination learning comes from studies comparing the effects of intradimensional (ID) and extradimensional (ED) shifts on performance. For instance, in an experiment reported by Mackintosh and Little (1969), pigeons were trained on two discrimination problems given in successive order. The stimuli were lines varying on the dimensions of colour and orientation, but the specific values of both dimensions were changed when the animals were shifted from the first to the second discrimination (“total change” design; Slamecka, 1968). In each discrimination problem, values from one relevant dimension (e.g., colour) signalled the availability of food, while values from the other dimension (e.g., orientation) were irrelevant to the solution of the discrimination (e.g., red lines were reinforced and yellow lines were nonreinforced, regardless of their orientation). Mackintosh and Little observed that the second discrimination was acquired more rapidly when it was based on a dimension already trained as relevant for the first discrimination (ID shift) than when it was based on a dimension previously trained as irrelevant (ED shift). The advantage in learning after an ID rather than an ED shift has been considered strong support for the assumption that discrimination training encourages organisms to pay more attention to stimuli relevant to the solution of a discrimination than to those that are irrelevant (see, e.g., Mackintosh, 1975, p. 279). ID–ED shift procedures have widespread applications in different fields of neuroscientific study. For instance, they have been used to investigate cognitive deficits in several clinical populations, including patients with frontal-lobe excisions (e.g., Owen, Roberts, Polkey, et al., 1991), Huntington’s disease

M. Uengoer (✉) · H. Lachnit
Faculty of Psychology, Philipps-Universität Marburg,
Gutenbergstraße 18,
35032 Marburg, Germany
e-mail: uengoer@staff.uni-marburg.de

(e.g., Lawrence, Sahakian, Rogers, et al., 1999), Parkinson's disease (e.g., Owen et al., 1993), or schizophrenia (e.g., Ceaser et al., 2008). ID–ED shift procedures have also been used to examine the functional properties of various neural circuits in nonhuman animals, including the prefrontal cortex in marmosets (e.g., Dias, Robbins, & Roberts, 1996), the medial frontal cortex in rats (e.g., Birrell & Brown, 2000), the entorhinal cortex in rats (e.g., Oswald et al., 2001), or the cingulate cortex in rats (e.g., Ng, Noblejas, Rodefer, et al. 2007).

In spite of the extensive use of ID–ED shift procedures, the factors guiding the modulation of attention in discrimination learning are still a matter of debate. At least two factors must be considered. First, the amount of attention paid to a stimulus might be guided by the correlation of this stimulus with the outcome. Consider, for instance, the discriminations trained in the experiment by Mackintosh and Little (1969). In each discrimination problem, values of the relevant dimension consistently signalled either reinforcement or nonreinforcement, whereas each value of the irrelevant dimension was associated with reward on 50% of the trials. Thus, it might be assumed that stimuli correlated with a specific outcome would receive more attention than those that are uncorrelated. Second, the amount of attention paid to a stimulus might be determined by the relevance of this stimulus for the solution of the discrimination. Stimuli providing the information required to solve the discrimination might receive more attention than stimuli making no contribution to the solution. To emphasise the difference between the correlation and relevance factors, consider a biconditional discrimination such as AB+, BC−, CD+, DA− (Saavedra, 1975). In this case, each individual stimulus is uncorrelated with the outcome, but nevertheless all stimuli are relevant to the solution of the discrimination, since the trial outcome is signalled by specific combinations of these stimuli.

In the majority of the studies using ID–ED shift procedures, the factors correlation and relevance are confounded (e.g., Birrell & Brown, 2000; Ceaser et al., 2008; Ng et al., 2007; Trobalon, Miguelez, McLaren, et al., 2003, to name just a few recent examples), making it impossible to assess their individual contributions to the modulation of attention in discrimination learning. A first step in disentangling this confound was taken by studies investigating the role of stimulus relevance using an experimental approach in which any potential contribution of the correlation factor was eliminated (George & Pearce, 1999; Kruschke, 1996; Oswald et al., 2001).

For instance, George and Pearce (1999) trained pigeons to discriminate between visual stimuli varying on three dimensions. Two of these dimensions were trained as relevant for the solution of a biconditional discrimination—that is, specific combinations of values of

these two dimensions signalled the trial outcome. The third dimension was irrelevant. Their design ensured that individual values of all three dimensions were equally correlated with the outcome (each value of each dimension was associated with a specific outcome on 50% of the trials), eliminating any potential contribution of this factor to the modulation of attention.

Subsequently, pigeons received a second discrimination for which the trial outcome was signalled by the values of a single dimension, and the remaining dimensions were irrelevant. George and Pearce (1999) found that the second discrimination was acquired more rapidly when it was based on a dimension already trained as relevant for the preceding biconditional discrimination than when it was based on a dimension previously trained as irrelevant. These results indicate that the amount of attention paid to a stimulus was controlled by the relevance of the stimulus and not solely by its correlation with the outcome (see also, for humans, Kruschke, 1996; for rats, Oswald et al., 2001). Thus, these studies identified that differences in stimulus relevance are sufficient to modulate attention in discrimination learning, and they demonstrated that differences in stimulus–outcome correlations are not necessary to encourage differences in attention. However, these studies have been silent about a potential contribution of the correlation factor to the modulation of attention in discrimination learning. To decide whether or not differences in stimulus–outcome correlations encourage differences in attention, it will be necessary, of course, to manipulate this factor explicitly (without any confounding manipulation of relevance).

The aim of the present experiments was to further disentangle the individual contributions of the factors correlation and relevance to the modulation of attention in discrimination learning. To assess the individual contribution of the factor correlation, we investigated in two experiments whether differences in stimulus–outcome correlations encouraged differences in attention, using an experimental design in which any potential contribution of the relevance factor was eliminated. To the best of our knowledge, the present study is the first to take this approach. In addition, both of our experiments aimed toward a further demonstration that differences in stimulus relevance are sufficient to encourage differences in attention (George & Pearce, 1999; Kruschke, 1996; Oswald et al., 2001). To demonstrate the generality of this finding, the present experiments implemented a “total change” design (Slamecka, 1968), which is a more common approach in studies using ID–ED shift procedures.

The present experiments examined modulation of attention in human discrimination learning using a predictive learning task. The experiments implemented a scenario in which participants were asked to imagine

being a special agent in a department for the investigation of serial crimes whose current case concerns a series of housebreakings in two boroughs of New York City: Brooklyn and Queens. Participants were told that at each crime scene the criminal leaves two cards announcing in which borough the next housebreaking will take place. The task was to find out via predictions and feedback which cards signalled a housebreak in Brooklyn and which cards signalled a housebreak in Queens. On successive trials, different stimulus compounds (each composed of two differently filled squares) were presented, each with a relation to a specific outcome (Brooklyn = “O1” or Queens = “O2”).

Experiment 1

Table 1 illustrates the design for the three groups of Experiment 1. Each group was trained with two successive discrimination problems. Each discrimination problem comprised four stimulus compounds, each composed of two elements from two dimensions. Each stimulus compound was paired with one of two different outcomes. In Phase 1, one group of participants (Group Irrelevant–Uncorrelated) received a linear discrimination, with one dimension (Dimension B) being relevant and another dimension (Dimension A) being irrelevant for the solution of the discrimination. Values of Dimension B signalled the trial outcome, whereas each value of Dimension A was followed by each of the outcomes on half of the trials ($B_1A_1 \rightarrow O1$, $B_1A_2 \rightarrow O1$, $B_2A_1 \rightarrow O2$, $B_2A_2 \rightarrow O2$; indices assign the values of a dimension). A second group of participants (Group Relevant–Correlated) were also trained with a linear discrimination, but with Dimension A being relevant and Dimension B being irrelevant

($A_1B_1 \rightarrow O1$, $A_1B_2 \rightarrow O1$, $A_2B_1 \rightarrow O2$, $A_2B_2 \rightarrow O2$). A third group of participants (Group Relevant–Uncorrelated) received a biconditional discrimination during Phase 1, with both Dimensions A and B being relevant: Specific combinations of values of both dimensions signalled the trial outcome, whereas each individual value of either dimension was followed by each of the outcomes on half of the trials ($A_1B_1 \rightarrow O1$, $A_1B_2 \rightarrow O2$, $A_2B_1 \rightarrow O2$, $A_2B_2 \rightarrow O1$). In Phase 2, all participants were trained with a linear discrimination in which the stimulus compounds were composed of new values of Dimension A and a new Dimension C. For the solution of this discrimination in each group, Dimension A was relevant and Dimension C irrelevant ($A_3C_1 \rightarrow O1$, $A_3C_2 \rightarrow O1$, $A_4C_1 \rightarrow O2$, $A_4C_2 \rightarrow O2$). Notice that group names indicate how the relevant Dimension A of Phase 2 had been treated in Phase 1. That is, in Group Irrelevant–Uncorrelated, the relevant dimension in Phase 2 had previously been trained as irrelevant for the discrimination and uncorrelated with the outcome. In Group Relevant–Correlated, the relevant dimension in Phase 2 had previously been trained as relevant and correlated. In Group Relevant–Uncorrelated, the relevant dimension in Phase 2 had previously been trained as relevant and uncorrelated.

The comparison between Group Irrelevant–Uncorrelated and Group Relevant–Uncorrelated provides information about the contribution of the factor stimulus relevance to the modulation of attention in discrimination learning. If stimuli that are relevant to the solution of a discrimination receive more attention than do stimuli that are irrelevant, the amount of attention paid to Dimension A in Phase 1 will be greater in Group Relevant–Uncorrelated than in Group Irrelevant–Uncorrelated. Therefore, at the outset of discrimination training in Phase 2, participants in Group Relevant–Uncorrelated will pay more attention to the relevant dimension than will the participants in Group Irrelevant–Uncorrelated. As a consequence, learning of the second discrimination should proceed faster in Group Relevant–Uncorrelated than in Group Irrelevant–Uncorrelated.

The comparison between Group Relevant–Correlated and Group Relevant–Uncorrelated allows for assessment of the contribution of the factor stimulus–outcome correlation. If stimuli that are correlated with the outcome receive more attention than do those that are uncorrelated, the amount of attention paid to Dimension A in Phase 1 will be greater in Group Relevant–Correlated than in Group Relevant–Uncorrelated. Therefore, at the outset of Phase 2, the relevant dimension will receive more attention in Group Relevant–Correlated than in Group Relevant–Uncorrelated. As a consequence, learning of the second discrimination should be facilitated in Group Relevant–Correlated as compared to Group Relevant–Uncorrelated.

Table 1 Design of Experiment 1

Phase 1	Phase 2
Group Irrelevant–Uncorrelated	
$B_1A_1 \rightarrow O1$, $B_1A_2 \rightarrow O1$	$A_3C_1 \rightarrow O1$, $A_3C_2 \rightarrow O1$
$B_2A_1 \rightarrow O2$, $B_2A_2 \rightarrow O2$	$A_4C_1 \rightarrow O2$, $A_4C_2 \rightarrow O2$
Group Relevant–Correlated	
$A_1B_1 \rightarrow O1$, $A_1B_2 \rightarrow O1$	$A_3C_1 \rightarrow O1$, $A_3C_2 \rightarrow O1$
$A_2B_1 \rightarrow O2$, $A_2B_2 \rightarrow O2$	$A_4C_1 \rightarrow O2$, $A_4C_2 \rightarrow O2$
Group Relevant–Uncorrelated	
$A_1B_1 \rightarrow O1$, $A_2B_2 \rightarrow O1$	$A_3C_1 \rightarrow O1$, $A_3C_2 \rightarrow O1$
$A_1B_2 \rightarrow O2$, $A_2B_1 \rightarrow O2$	$A_4C_1 \rightarrow O2$, $A_4C_2 \rightarrow O2$

A–C are different dimensions and the respective indices assign specific values of a dimension; O1 = “Brooklyn”; O2 = “Queens.”

Method

Participants The participants were 36 students of Philipps-Universität Marburg (17 females, 19 males). Their age varied between 19 and 33 years, with a median of 23. They either participated in order to meet course requirements or were paid with sweets. Participants were randomly allocated to the different experimental groups as they arrived at the experimental room. They were tested individually and required approximately 15 min to complete the experiment. For 4 additional participants, data collection was aborted because they did not reach the learning criterion for Phase 1 (see below) within 160 trials.

Apparatus and procedure The instructions and all necessary information were presented on a computer screen. Participants interacted with the computer using the mouse. A total of 16 differently filled squares (with a side length of 4 cm each) were used as cues. Each of four of these squares belonged to one of the four stimulus types (colour, orientation, shape, and letter). Squares of the stimulus type Colour were each filled with a solid colour (green, red, blue, or yellow). Squares of the stimulus type Orientation were occupied by alternating black and white 2-mm-wide stripes displayed in one of four orientations (45°, 135°, horizontal, or vertical). Each square of the stimulus type Shape displayed a white line drawing of one of four geometric forms (circle, parallelogram, star, or triangle) on a black background. Each square of stimulus type Letter showed one of four capital letters (G, K, M, or S) in black font on a white background. In each group, Dimension A was realised by stimulus type Orientation. The assignment of the remaining stimulus types Colour, Shape, and Letter to Dimensions B and C was counterbalanced across participants within each group. The assignment of the members of one stimulus type to the values of a dimension was implemented randomly for each participant, with one restriction: Either the 45° stimulus and the 135° stimulus or the horizontal stimulus and the vertical stimulus were allocated together to the same learning phase. The two different outcomes were the names of two boroughs of New York City, “Brooklyn” (O1) and “Queens” (O2).

Each participant was initially asked to read the following instructions (in German) on the screen:

This study is concerned with the question of how people learn about relationships between different events. Imagine that you are a special agent in a department for the investigation of serial crimes and that the New York Police Department seeks your support. The current case is about a series of mysterious housebreakings in two boroughs of New York City, Brooklyn and Queens. The special thing

about this case is that at each crime scene the criminal leaves two cards. Your colleagues suppose that by these cards left at a crime scene the criminal announces in which borough he will strike next. Your job now is to pursue this lead. Therefore, you receive access to the files of the inquiry. In the following, you will search the files of the housebreakings in a chronological order. For each housebreaking, you will be shown which cards the criminal left at the crime scene. Thereafter, you will be asked to predict whether the next housebreaking took place in Brooklyn or in Queens. For this prediction, there will be two appropriate response buttons available. After you have made your prediction, you will be informed in which borough the next housebreaking actually took place. Use this feedback to find out which cards signal a housebreak in Brooklyn and which cards signal a housebreak in Queens.

Obviously, at first you will have to guess, as you do not know anything about the criminal. But eventually you will learn about the method by which the criminal acts. On the basis of this knowledge, you should make correct predictions—as many as possible. For all of your answers, accuracy rather than speed is essential. Please do not take any notes during the experiment. If you have any more questions please ask them now. If you don’t have any questions, please start the experiment by clicking on the “Next” button.

When a participant asked a question, it was answered by the experimenter. After clicking on a button labelled “Next,” the learning phases started.

On each learning trial, a stimulus compound composed of two squares from different stimulus types was shown on the top half of the screen. The two squares were presented side by side, with the left–right allocation determined randomly on each trial. Each square appeared at a distance of 4 cm from the vertical centre of the display. Participants were told that the squares were cards left by the criminal at the crime scene. They were also asked to predict whether the next housebreaking would take place in Brooklyn or Queens. Participants made their predictions by clicking on one of two answer buttons labelled “Brooklyn” or “Queens.” Immediately after they responded, another window appeared, telling the participants in which of the two boroughs the next housebreaking had actually taken place. Participants had to confirm that they had read the feedback by clicking on an “OK” button. Thereafter, the next trial started.

Three groups of participants each worked on two learning phases. During the first learning phase, one group of participants (Group Irrelevant–Uncorrelated) was trained with a linear discrimination with Dimension B being

relevant and Dimension A being irrelevant for its solution ($B_1A_1 \rightarrow O1$, $B_1A_2 \rightarrow O1$, $B_2A_1 \rightarrow O2$, $B_2A_2 \rightarrow O2$). A second group of participants (Group Relevant–Correlated) was also trained with a linear discrimination, but with Dimension A being relevant and Dimension B being irrelevant ($A_1B_1 \rightarrow O1$, $A_1B_2 \rightarrow O1$, $A_2B_1 \rightarrow O2$, $A_2B_2 \rightarrow O2$). A third group of participants (Group Relevant–Uncorrelated) received a biconditional discrimination with Dimensions A and B being relevant ($A_1B_1 \rightarrow O1$, $A_1B_2 \rightarrow O2$, $A_2B_1 \rightarrow O2$, $A_2B_2 \rightarrow O1$). In Phase 2, all participants were trained with a linear discrimination with Dimension A being relevant and Dimension C being irrelevant ($A_3C_1 \rightarrow O1$, $A_3C_2 \rightarrow O1$, $A_4C_1 \rightarrow O2$, $A_4C_2 \rightarrow O2$).

For each group, each learning phase was divided into blocks. Within each block, each trial type was presented on two occasions. The order of presentation of the trials within each block was determined randomly for each block and each participant, with one restriction: No more than three stimulus compounds followed each other containing the same stimulus element. For each group, Phase 1 consisted of at least two blocks. After the second block, the number of further blocks given to a participant depended on his or her prediction accuracy. If the participant's predictions had been correct on at least 80% of the trials during the last two blocks (learning criterion), the second learning phase started (the transition from Phase 1 to Phase 2 was not signalled to the participants); otherwise, Phase 1 was extended for a further block. Phase 2 consisted of five blocks for all participants.

Data scoring and aggregation For each participant, we calculated the proportion of correct predictions for each block of eight trials in Phases 1 and 2. We also recorded for each participant the number of trials required to complete the training in Phase 1.

Results and discussion

For this and the subsequent experiment, the .05 level of significance was employed for all statistical tests, and the stated probability levels are based on the Greenhouse and Geisser (1959) adjustment of degrees of freedom, where appropriate.

Learning in Phase 1 To assess performance in Phase 1, we compared between groups the mean proportions of correct predictions collapsed across the last two blocks of Phase 1 using a one-way ANOVA. This analysis revealed no difference between groups (Group Irrelevant–Uncorrelated, $M = 88\%$; Group Relevant–Correlated, $M = 89\%$; Group Relevant–Uncorrelated, $M = 88.5\%$), $F < 1$. In addition, a

one-way ANOVA between groups, comparing the number of trials required to reach the learning criterion in Phase 1, yielded a significant effect of group, $F(2, 35) = 6.14$, $p < .01$. Post-hoc tests using the Bonferroni correction revealed that participants in Group Relevant–Uncorrelated ($M = 54.7$) required more trials than did participants in Group Irrelevant–Uncorrelated ($M = 23.3$), $p < .02$, or participants in Group Relevant–Correlated ($M = 25.3$), $p < .02$, whereas participants in the latter two groups did not differ in their required numbers of trials, $p = 1.0$. Taken together, the linear discriminations trained in Groups Irrelevant–Uncorrelated and Relevant–Correlated were learned within fewer trials than was the biconditional discrimination trained in Group Relevant–Uncorrelated, but at the end of Phase 1, all groups were equated for accuracy.

Learning in Phase 2 Figure 1 presents the mean proportions of correct predictions across the five blocks of Phase 2 for each group. Black squares represent data from Group Irrelevant–Uncorrelated, grey triangles data from Group Relevant–Correlated, and white circles data from Group Relevant–Uncorrelated.

To assess performance in Phase 2, we analysed the proportions of correct predictions using a 5×3 repeated measures ANOVA, including the within-subjects factor block (1–5) and the between-subjects factor group (Irrelevant–Uncorrelated vs. Relevant–Correlated vs. Relevant–Uncorrelated). The analysis revealed a main effect of block, $F(4, 132) = 8.20$, $p < .001$, indicating that the accuracy of predictions increased across the five blocks of Phase 2. There was also a main effect of group, $F(1, 33) = 3.60$, $p < .04$, reflecting that the accuracy of predictions differed between

Fig. 1 Mean proportions of correct predictions across the five blocks of Phase 2 of Experiment 1. Black squares correspond to Group Irrelevant–Uncorrelated, grey triangles to Group Relevant–Correlated, and white circles to Group Relevant–Uncorrelated

groups. The block $\times$ group interaction was not significant, $F < 1$. Using Fisher's least significant difference (LSD) test, we conducted post-hoc comparisons between groups analysing the mean proportions of correct predictions, collapsed across the five blocks of Phase 2. These analyses showed that the accuracy of the predictions in Group Relevant–Uncorrelated ($M = 82.7\%$) was higher when compared with that in Group Irrelevant–Uncorrelated ($M = 65.4\%$), $p < .04$, but was equal to the accuracy in Group Relevant–Correlated ($M = 84.2\%$), $p > .85$.

Overall, our participants acquired the second discrimination more rapidly when it was based on a dimension previously trained as relevant and uncorrelated than when it was based on a dimension previously trained as irrelevant and uncorrelated. This result indicates that differences in stimulus relevance are sufficient to encourage differences in attention. Although this finding had been documented in previous experiments by George and Pearce (1999), Kruschke (1996), and Oswald et al. (2001), the present experiment is the first to show this effect in a “total change” design (Slamecka, 1968). In addition, we found that the rate of acquisition of the second discrimination was independent of whether the relevant dimension had previously been trained as relevant and uncorrelated or as relevant and correlated. This finding supports the view that differences in stimulus–outcome correlations are not sufficient to encourage differences in attention. Thus, the results from the present experiment suggest that the modulation of attention in discrimination learning is guided by stimulus relevance and not by stimulus–outcome correlation.

In the present experiment, we manipulated the relevant dimension of the second discrimination across groups during Phase 1, whereas the dimension trained as irrelevant for the second discrimination was equally novel for all groups. To enhance the generality of the conclusions drawn from the present experiment, it was necessary, however, to show that these conclusions were independent of the specific test strategy used. Therefore, Experiment 2 was designed to provide another test of the proposal that the modulation of attention in discrimination learning is guided by stimulus relevance and not by stimulus–outcome correlation. For this purpose, Experiment 2 differed from Experiment 1 in that the irrelevant dimension of the second discrimination was manipulated across groups during Phase 1, whereas the dimension trained as relevant for the second discrimination was novel for all groups.

Experiment 2

The design of Experiment 2 is shown in Table 2. Each of three groups of participants was trained with two successive

Table 2 Design of Experiment 2

Phase 1	Phase 2
Group Irrelevant–Uncorrelated	
$B_1A_1 \rightarrow O1$, $B_1A_2 \rightarrow O1$	$C_1A_3 \rightarrow O1$, $C_1A_4 \rightarrow O1$
$B_2A_1 \rightarrow O2$, $B_2A_2 \rightarrow O2$	$C_2A_3 \rightarrow O2$, $C_2A_4 \rightarrow O2$
Group Relevant–Correlated	
$A_1B_1 \rightarrow O1$, $A_1B_2 \rightarrow O1$	$C_1A_3 \rightarrow O1$, $C_1A_4 \rightarrow O1$
$A_2B_1 \rightarrow O2$, $A_2B_2 \rightarrow O2$	$C_2A_3 \rightarrow O2$, $C_2A_4 \rightarrow O2$
Group Relevant–Uncorrelated	
$A_1B_1 \rightarrow O1$, $A_2B_2 \rightarrow O1$	$C_1A_3 \rightarrow O1$, $C_1A_4 \rightarrow O1$
$A_1B_2 \rightarrow O2$, $A_2B_1 \rightarrow O2$	$C_2A_3 \rightarrow O2$, $C_2A_4 \rightarrow O2$

A–C are different dimensions and the respective indices assign specific values of a dimension; O1 = “Brooklyn”; O2 = “Queens.”

discrimination problems. Phase 1 of Experiment 2 was identical to Phase 1 of Experiment 1. That is, one group (Group Irrelevant–Uncorrelated) received a linear discrimination with Dimension B being relevant and Dimension A being irrelevant ($B_1A_1 \rightarrow O1$, $B_1A_2 \rightarrow O1$, $B_2A_1 \rightarrow O2$, $B_2A_2 \rightarrow O2$), a second group (Group Relevant–Correlated) worked on a linear discrimination with Dimension A being relevant and Dimension B being irrelevant ($A_1B_1 \rightarrow O1$, $A_1B_2 \rightarrow O1$, $A_2B_1 \rightarrow O2$, $A_2B_2 \rightarrow O2$), and a third group (Group Relevant–Uncorrelated) received a biconditional discrimination, with Dimensions A and B both being relevant ($A_1B_1 \rightarrow O1$, $A_1B_2 \rightarrow O2$, $A_2B_1 \rightarrow O2$, $A_2B_2 \rightarrow O1$). In Phase 2, departing from Experiment 1, all participants were trained with a linear discrimination problem, with Dimension C being relevant and Dimension A being irrelevant ($C_1A_3 \rightarrow O1$, $C_1A_4 \rightarrow O1$, $C_2A_3 \rightarrow O2$, $C_2A_4 \rightarrow O2$). Notice that group names indicate how the irrelevant Dimension A of Phase 2 was treated in Phase 1. That is, in Group Irrelevant–Uncorrelated, the dimension that was irrelevant in Phase 2 had previously been trained as irrelevant and uncorrelated. In Group Relevant–Correlated, the dimension that was irrelevant in Phase 2 had previously been trained as relevant and correlated. In Group Relevant–Uncorrelated, the dimension that was irrelevant in Phase 2 had previously been trained as relevant and uncorrelated.

If the amount of attention paid to a stimulus is determined by the relevance of this stimulus for the solution of a discrimination, those stimuli that are irrelevant in Phase 2 would receive more attention in Group Relevant–Uncorrelated than in Group Irrelevant–Uncorrelated. Therefore, learning of the second discrimination should proceed at a lower rate in Group Relevant–Uncorrelated than in Group Irrelevant–Uncorrelated. If attention to a stimulus is guided by the correlation of this stimulus with the outcome, those stimuli that are irrelevant in Phase 2 will receive more attention in Group Relevant–

Correlated than in Group Relevant–Uncorrelated, and therefore, acquisition of the discrimination in Phase 2 should be retarded in Group Relevant–Correlated as compared to in Group Relevant–Uncorrelated.

Method

Participants A group of 36 students of Philipps-Universität Marburg (28 females, 8 males) participated in [Experiment 2](#). Their age varied between 19 and 27 years, with a median of 21. They either participated in order to meet course requirements or were paid with sweets. Participants were randomly allocated to the different experimental groups as they arrived at the experimental room. They were tested individually and required approximately 15 min to complete the experiment. For 3 additional participants, data collection was broken off because they did not reach the learning criterion for Phase 1 (see the [Method](#) section of [Experiment 1](#)) within 160 trials.

Apparatus, procedure, data scoring, and aggregation The apparatus, procedure, data scoring, and data aggregation were identical to those aspects of [Experiment 1](#), unless stated otherwise. In Phase 2, all participants were trained with a linear discrimination, with Dimension C being relevant and Dimension A being irrelevant ($C_1A_3 \rightarrow O1$, $C_1A_4 \rightarrow O1$, $C_2A_3 \rightarrow O2$, $C_2A_4 \rightarrow O2$).

Results and discussion

Learning in Phase 1 A one-way ANOVA revealed no difference between groups in the mean proportions of correct predictions collapsed across the last two blocks of Phase 1 (Group Irrelevant–Uncorrelated, $M = 91.2\%$; Group Relevant–Correlated, $M = 89.1\%$; Group Relevant–Uncorrelated, $M = 87\%$), $F(2, 35) = 1.16$, $p > .32$. A one-way ANOVA between groups comparing the numbers of trials required to reach the learning criterion in Phase 1 yielded a significant effect of group, $F(2, 35) = 8.72$, $p < .001$. Post-hoc tests using the Bonferroni correction revealed that participants in Group Relevant–Uncorrelated ($M = 47.3$) required more trials than did participants in Group Irrelevant–Uncorrelated ($M = 22.7$), $p < .01$, or participants in Group Relevant–Correlated ($M = 24$), $p < .01$, whereas participants in Groups Irrelevant–Uncorrelated and Relevant–Correlated did not differ in their required numbers of trials, $p = 1.0$. Thus, as in [Experiment 1](#), the linear discriminations trained in Group Irrelevant–Uncorrelated and Group Relevant–Correlated were learned within fewer trials than was the biconditional discrimination trained in Group Relevant–Uncorrelated, but at the end of Phase 1, all groups were equated for accuracy.

Learning in Phase 2 Figure 2 presents the mean proportions of correct predictions across the five blocks of Phase 2 for each group. Black squares represent the data from Group Irrelevant–Uncorrelated, grey triangles data from Group Relevant–Correlated, and white circles data from Group Relevant–Uncorrelated.

To assess performance in Phase 2, we analysed the proportions of correct predictions using a block (1–5) $\times$ group (Irrelevant–Uncorrelated vs. Relevant–Correlated vs. Relevant–Uncorrelated) ANOVA. The analysis revealed a main effect of block, $F(4, 132) = 19.50$, $p < .001$, indicating that the accuracy of predictions increased across the five blocks of Phase 2, and a significant block $\times$ group interaction, $F(8, 132) = 2.48$, $p < .05$, showing that the increase in prediction accuracy proceeded differently between groups. The main effect of group was not significant, $F(2, 33) = 1.80$, $p > .18$. To decompose the block $\times$ group interaction, we calculated simple main effects of group at each level of the block factor. These analyses revealed a simple main effect of group for Block 1, $F(2, 33) = 3.78$, $p < .04$, showing that at this level prediction accuracy differed between the groups. For the remaining blocks, there were no simple main effects of group, $F_s < 1.04$, $p_s > .36$. Using Fisher's LSD test, we conducted post-hoc comparisons between groups analysing the proportions of correct predictions for Block 1. These analyses yielded that the accuracy of predictions was lower in Group Relevant–Uncorrelated ($M = 77.1\%$) than in Group Irrelevant–Uncorrelated ($M = 93.8\%$), $p < .02$, but equal to the accuracy in Group Relevant–Correlated ($M = 79.2\%$), $p > .75$.

Fig. 2 Mean proportions of correct predictions across the five blocks of Phase 2 of [Experiment 2](#). Black squares correspond to Group Irrelevant–Uncorrelated, grey triangles to Group Relevant–Correlated, and white circles to Group Relevant–Uncorrelated

Overall, acquisition of the second discrimination proceeded at a lower rate when its irrelevant dimension had previously been trained as relevant and uncorrelated than when its irrelevant dimension had previously been trained as irrelevant and uncorrelated. We also found that the rates of acquisition of the second discrimination were independent of whether the irrelevant dimension had previously been trained as relevant and uncorrelated or as relevant and correlated. Thus, the results from [Experiment 2](#) reveal converging evidence that the modulation of attention in discrimination learning is guided by stimulus relevance, and that differences in stimulus–outcome correlations are not sufficient to encourage differences in attention.

General discussion

In two experiments, we investigated the importance of stimulus relevance and stimulus–outcome correlation to the modulation of attention in discrimination learning. We found that acquisition of a discrimination was influenced by whether its relevant dimension ([Experiment 1](#)) or its irrelevant dimension ([Experiment 2](#)) had previously been trained as relevant and uncorrelated or as irrelevant and uncorrelated. We also observed that acquisition of a discrimination was independent of whether its relevant dimension ([Experiment 1](#)) or its irrelevant dimension ([Experiment 2](#)) had previously been trained as relevant and uncorrelated or as relevant and correlated. This pattern of results indicates that the amount of attention paid to a stimulus during discrimination learning is determined by the relevance of this stimulus for the solution of the discrimination, and not by its correlation with the outcome.

Our study demonstrates the generality of the conclusion drawn from previous experiments that differences in stimulus relevance are sufficient to encourage differences in attention (George & Pearce, 1999; Kruschke, 1996; Oswald et al., 2001). Our study is the first to provide evidence for this conclusion in a “total change” design (Slamecka, 1968). In addition, the present study also complements the previous studies from George and Pearce (1999), Kruschke (1996), and Oswald et al. (2001) by demonstrating that differences in stimulus–outcome correlations are not sufficient to encourage differences in attention.

Before turning to a discussion of our results in terms of attentional changes in the framework of theories of associative learning and attention, we need to consider an alternative explanation that would require no recourse to attentional processes. Assume, for instance, that the participants in each of our experiments started Phase 2 with certain rules that they had derived from their

experiences in Phase 1. In Group Irrelevant–Uncorrelated, our participants might have started the second discrimination with the rule that Dimension A was again irrelevant (and the alternative dimension was relevant). In Group Relevant–Correlated, participants might have started Phase 2 with the rule that Dimension A was again relevant. And in Group Relevant–Uncorrelated, participants might have assumed that the second task was again a biconditional discrimination. On the first trial in Phase 2, participants would still have had to guess when asked to predict the correct outcome, but on the basis of their rules, participants might draw conclusions from the feedback given on that trial about the remaining stimulus–outcome relations.

This inferential account was able to deal with our results from [Experiment 2](#). The approach predicts for Group Irrelevant–Uncorrelated in [Experiment 2](#) that, on the basis of the feedback given on Trial 1 of Phase 2, participants would be able to correctly infer the remaining stimulus–outcome relations. The Trial 1 response could well be an error, but thereafter, performance could be perfect. The rules postulated for the other two groups would require guessing on trials following Trial 1 in order for these participants to infer the remaining stimulus–outcome relations. Therefore, this approach correctly predicted that the rate of acquisition of the second discrimination in [Experiment 2](#) would be higher in Group Irrelevant–Uncorrelated than in Group Relevant–Uncorrelated, but equal between Groups Relevant–Uncorrelated and Relevant–Correlated. In addition, this approach was able to deal with the high level of accuracy in Group Irrelevant–Uncorrelated (93%) observed in Block 1 of Phase 2 of [Experiment 2](#), indicating that on average, participants made only 0.5 errors during this block.

However, if the same principles advocated by the inferential account were applied to [Experiment 1](#), the approach would lead to several incorrect predictions. The approach assumes for Group Relevant–Correlated in [Experiment 1](#) that on the basis of the feedback given on Trial 1 of Phase 2, participants would be able, using their rule, to correctly infer the remaining stimulus–outcome relations. In each of the remaining groups, participants would make two errors if they applied their rule to the feedback on Trial 1 to infer the remaining stimulus–outcome relations. Thus, this approach incorrectly predicted that the rate of acquisition of the second discrimination in [Experiment 1](#) should have been higher in Group Relevant–Correlated than in Group Relevant–Uncorrelated, but equal between Groups Relevant–Uncorrelated and Irrelevant–Uncorrelated. Neither prediction was supported by the results of [Experiment 1](#). In addition, Group Relevant–Correlated showed a level of accuracy of 75% in Block 1 of Phase 2 of [Experiment 1](#). It is hard to reconcile this observed level of accuracy with the assumption that after

the first trial in Phase 2, participants in this group should have been able to correctly infer the stimulus–outcome relations trained during this phase. We now explore whether theories of associative learning and attention can be used to explain the results from Experiments 1 and 2.

The idea that attentional changes are governed by stimulus–outcome correlations is central to the theory of attention proposed by Mackintosh (1975). According to this theory, attention to a stimulus will increase if an outcome is predicted more accurately on the basis of this stimulus than on the basis of all other stimuli concurrently present, whereas attention to a stimulus will decrease if an outcome is predicted more accurately by other accompanying stimuli. The theory of Mackintosh adopts a purely elemental stimulus representation assuming that each element of a stimulus compound acquires its own direct association with the outcome. Therefore, this theory cannot account for the acquisition of a biconditional discrimination, as observed in each of our experiments. One way to resolve this problem would be to adopt the unique-cue hypothesis (see, e.g., Wagner & Rescorla, 1972), which supposes that a compound consists of its elements plus an additional element unique to the specific stimulus conjunction. Learning about these unique cues is thought to proceed according to the same rules as learning of any other stimulus.

George and Pearce (1999), for instance, suggested a unique-cue extension of Mackintosh's (1975) theory, in which both increases and decreases in attention would be allowed to generalise between unique cues and their constituting elements. In order to derive precise predictions from this approach for the present experiments, however, it will be necessary to further specify some of the properties of such a model. First, we assume that changes in attention can generalise from three different sources: generalisation between stimuli belonging to the same dimension (${}_AG_A$), generalisation between unique cues and their elements (${}_AG_{AB}$), and generalisation between unique cues sharing a common element (${}_{AB}G_{AC}$). Second, we assume that the degree of generalisation of attentional changes is a function of similarity, so that ${}_AG_A$, ${}_AG_{AB}$, and ${}_{AB}G_{AC}$ differ in effectiveness, with ${}_AG_{AB} = 0.5{}_AG_A$ and ${}_{AB}G_{AC} = 0.5{}_AG_{AB}$. Third, we assume low salience for unique cues, so that they only acquire substantial associative strength if they are required for successful performance, as it is in learning of a biconditional discrimination. Furthermore, we set the starting value of the attention parameter α to .5 and allowed α to vary between 0 and 1.

Under these conditions, a unique-cue extension of Mackintosh's (1975) theory is able to deal with our results from Experiment 1. The model predicts for the first phase of our experiments that in Groups Irrelevant–Uncorrelated and Relevant–Correlated, attention to each of the two

stimuli belonging to the relevant dimension will increase to 1, whereas attention to each of the two stimuli belonging to the irrelevant dimension and to each of the four unique cues will decrease to 0. In Group Relevant–Uncorrelated, attention to each unique cue will increase to 1, whereas attention to each stimulus belonging to Dimension A or Dimension B will decrease to 0. Now consider how these attentional changes generalise to the stimuli belonging to Dimension A in Phase 2 of our experiments. In Group Irrelevant–Uncorrelated, each stimulus from Dimension A in Phase 2 receives a generalised decrease in attention of $2{}_AG_A$ from the stimuli belonging to Dimension A and a generalised decrease in attention of $4{}_AG_{AB}$ from the unique cues. In Group Relevant–Correlated, each stimulus from Dimension A in Phase 2 receives a generalised increase in attention of $2{}_AG_A$ from the stimuli belonging to Dimension A, but this increase in attention is cancelled by a generalised decrease in attention of $4{}_AG_{AB}$ from the unique cues. In Group Relevant–Uncorrelated, each stimulus from Dimension A in Phase 2 receives a generalised decrease in attention of $2{}_AG_A$ from the stimuli belonging to Dimension A, but this decrease in attention is cancelled by a generalised increase in attention of $4{}_AG_{AB}$ from the unique cues. Thus, at the outset of Phase 2, attention to the stimuli belonging to Dimension A will be higher in Group Relevant–Uncorrelated than in Group Irrelevant–Uncorrelated, but equal between Groups Relevant–Uncorrelated and Relevant–Correlated. For Experiment 1, in which Dimension A was relevant for the solution of the second discrimination, the model correctly predicts the pattern of group differences we observed during Phase 2. However, for Experiment 2, in which Dimension A was irrelevant in Phase 2, the model erroneously predicts that the rates of acquisition of the second discrimination should have been equal across all groups. This prediction seems to arise from the use of a separable error term in Mackintosh's theory, ensuring that associative learning about the stimuli belonging to the relevant Dimension C can proceed unimpeded from concurrently presented stimuli.

Another theoretical framework for the discussion of our experiments is provided by a connectionist model of category learning named ALCOVE (Kruschke, 1992). If certain assumptions are made, ALCOVE is able to account for the present pattern of results. ALCOVE is an exemplar-based three-layer network. The model assumes a stimulus representation that characterises stimuli as points in a multidimensional psychological space. The input layer consists of nodes, each corresponding to a single psychological dimension, for which the activation of a specific node represents the value of a given stimulus on that particular dimension. The output layer comprises nodes representing response categories. The input and output layers are interconnected via an intermediary hidden layer. Each hidden

node represents a training exemplar encoded as a point in the multidimensional stimulus space. The activation of a specific hidden node depends on the similarity between the exemplar represented by this node and the external stimulus.

The functional role of attention in this model is to increase or decrease the importance of individual dimensions for the calculation of this similarity between an exemplar and a stimulus. Therefore, each input node is related to a dimensional attention strength, α_i , adjusted to minimise the error function of the network (Rumelhart, Hinton, & Williams, 1986). With increasing attention strength, its corresponding dimension is weighted stronger in the computation of the metric distance between exemplars and stimuli. To illustrate the consequences of this weighted similarity computation, consider, as an example, two stimuli differing in shape but with the same colour (blue square and blue triangle). If the attention strength on the dimension shape is high and the attention strength on the dimension colour is low, both stimuli will activate different hidden nodes, which will facilitate discrimination learning between these stimuli. However, if the attention strength on the dimension shape is low and the attention strength on the dimension colour is high, both stimuli will activate the same hidden nodes, which will interfere with discrimination learning. By default, learning within the ALCOVE model starts with equal attention strengths on all dimensions. In the course of discrimination training, the model increases the attention strengths related to dimensions relevant for the solution of the discrimination, and decreases the attention strengths corresponding to those that are irrelevant to the task. Whether or not ALCOVE can deal with the present results depends on the constraints subjected to the attention strengths. In its original formulation, attention strengths were only constrained to be nonnegative, $\alpha_i \geq 0$ (Kruschke, 1992, p. 24). Under this condition, ALCOVE predicts for the present experiments that the attention strength on a relevant and correlated dimension trained in a linear discrimination, $\alpha_{\text{Rel-Cor}}$, will be higher than the attention strength on an irrelevant and uncorrelated dimension trained in a linear discrimination, $\alpha_{\text{Irrel-Uncor}}$, but lower than the attention strength on a relevant and uncorrelated dimension trained in a biconditional discrimination, $\alpha_{\text{Rel-Uncor}}$ ($\alpha_{\text{Irrel-Uncor}} < \alpha_{\text{Rel-Cor}} < \alpha_{\text{Rel-Uncor}}$). Hence, ALCOVE correctly anticipates our observation that Groups Relevant–Uncorrelated and Irrelevant–Uncorrelated acquired the second discrimination at different rates. However, this setting also leads ALCOVE to erroneously predict a difference between the rates at which Groups Relevant–Uncorrelated and Relevant–Correlated would acquire the second discrimination. One way to overcome this problem would be to restrict the processing capacity of ALCOVE by limiting possible increases of each attention strength to an upper boundary (e.g., $1 \geq \alpha_i \geq 0$).

Under this condition, the attention strength on a relevant and correlated dimension would be equal to the attention strength on a relevant and uncorrelated dimension ($\alpha_{\text{Irrel-Uncor}} < \alpha_{\text{Rel-Cor}} = \alpha_{\text{Rel-Uncor}}$), which is consistent with the pattern of results from each of our experiments.

Our results indicate that the modulation of attention is guided by stimulus relevance and not by stimulus–outcome correlation. However, each of our experiments is silent about the way in which differences in attention between relevant and irrelevant stimuli might arise. At least two possibilities must be considered. First, differences in attention might arise from increases in attention to relevant stimuli and decreases in attention to irrelevant stimuli. Second, differences in attention might arise mainly from decreases in attention to irrelevant stimuli. Our experiments were not designed to decide between these two possibilities, but cross-experimental comparisons might allow us to draw conclusions concerning this issue. If differences in attention arise equally from increases in attention to relevant and decreases in attention to irrelevant stimuli, then the acquisition of the second discrimination should be equally demanding for participants in Experiments 1 and 2. However, if differences in attention arise more from decreases in attention to irrelevant stimuli than from increases in attention to relevant stimuli, acquisition of the second discrimination should be easier for participants in Experiment 2 than in Experiment 1. A cross-experimental comparison rather supports the latter possibility. We found that the proportion of correct predictions for Block 1 in Phase 2 was higher in Experiment 2 (83%) than in Experiment 1 (67%), $t(70) = -3.17$, $p < .01$. However, conclusions from cross-experimental comparisons should be treated with caution, so future research will be required in order to further specify the dynamics of the attentional changes in discrimination learning.

Author note The research reported in this article was supported by Grant 564/19-1 from the German Research Foundation (Deutsche Forschungsgemeinschaft) to H.L. We thank John K. Kruschke for providing us with a simulation program of his ALCOVE model. We also thank Jana Becht, Veronika Hartl, Sarah Keller, Liridon Koreaj, Lynn Antje Kovarik, Pit Marquardt, and Farsin Siegmund for their help with data collection. The first author of this paper has previously published with “Üngör” as his surname.

References

- Birrell, J. M., & Brown, V. J. (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. *Journal of Neuroscience*, 20, 4320–4324.
- Ceaser, A. E., Goldberg, T. E., Egan, M. F., McMahon, R. P., Weinberger, D. R., & Gold, J. M. (2008). Set-shifting ability and schizophrenia: A marker of clinical illness or an intermediate phenotype? *Biological Psychiatry*, 64, 782–788. doi:10.1016/j.biopsych.2008.05.009
- Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature*, 380, 69–72.

- 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.
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112. doi:10.1007/BF02289823
- Hall, G. (1991). *Perceptual and associative learning*. Oxford: Oxford University Press.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44. doi:10.1037/0033-295X.99.1.22
- Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science*, 8, 225–247. doi:10.1080/095400996116893
- Lawrence, A. D., Sahakian, B. J., Rogers, R. D., Hodges, J. R., & Robbins, T. W. (1999). Discrimination, reversal, and shift learning in Huntington's disease: Mechanisms of impaired response selection. *Neuropsychologia*, 37, 1359–1374.
- Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology*, 57B, 193–243.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276–298. doi:10.1037/h0076778
- Mackintosh, N. J., & Little, L. (1969). Intradimensional and extradimensional shift learning by pigeons. *Psychonomic Science*, 14, 5–6.
- Ng, C. W., Noblejas, M. I., Rodefer, J. S., Smith, C. B., & Poremba, A. (2007). Double dissociation of attentional resources: Prefrontal versus cingulate cortices. *Journal of Neuroscience*, 27, 12123–12131.
- Oswald, C. J. P., Yee, B. K., Rawlins, J. N. P., Bannerman, D. B., Good, M., & Honey, R. C. (2001). Involvement of the entorhinal cortex in a process of attentional modulation: Evidence from a novel variant of an IDS/EDS procedure. *Behavioral Neuroscience*, 115, 841–849.
- Owen, A. M., Roberts, A. C., Hodges, J. R., Summers, B. A., Polkey, C. E., & Robbins, T. W. (1993). Contrasting mechanisms of impaired attentional set-shifting in patients with frontal lobe damage or Parkinson's disease. *Brain*, 116, 1159–1175.
- Owen, A. M., Roberts, A. C., Polkey, C. E., Sahakian, B. J., & Robbins, T. W. (1991). Extra-dimensional versus intra-dimensional set shifting performance following frontal-lobe excisions, temporal-lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia*, 29, 993–1006.
- 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.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing* (Vol. 1, chap. 8). Cambridge, MA: MIT Press.
- Saavedra, M. A. (1975). Pavlovian compound conditioning in the rabbit. *Learning and Motivation*, 6, 314–326.
- Slamecka, N. J. (1968). A methodological analysis of shift paradigms in human discrimination learning. *Psychological Bulletin*, 69, 423–438.
- Trobalon, J. B., Miguelez, D., McLaren, I. P. L., & Mackintosh, N. J. (2003). Intradimensional and extradimensional shifts in spatial learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 29, 143–152.
- 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.