Journal of Experimental Psychology: Animal Learning and Cognition
2016, Vol. 42, No. 1, 15–31

2329-8456/16/$12.00 http://dx.doi.org/10.1037/xan0000082

Transfer of Dimensional Associability in Human Contingency Learning

Florian Kattner and C. Shawn Green
University of Wisconsin–Madison

Several studies have demonstrated processing advantages for stimuli that were experienced to be reliable predictors of an outcome relative to other stimuli. The present study tested whether such increases in associability apply at the level of entire stimulus dimensions (as suggested by Sutherland & Mackintosh, 1971). In 4 experiments, participants had to learn associations between Gabor gratings and particular responses. In a first experiment, some gratings were more predictive of the response than other gratings, whereas in 3 subsequent experiments, one stimulus dimension (i.e., either the orientation or spatial frequency of the grating) was more predictive than the other dimension. In contrast to the learned predictiveness of individual gratings (Experiment 1), dimensional predictiveness did not affect the subsequent rate of learning (Experiments 2 and 3), suggesting changes in the associability of specific stimuli, but not of stimulus dimensions. Moreover, greater transfer of predictiveness was found in all experiments when particular stimulus values of the test discrimination did not lie between the previously relevant stimuli. In Experiment 4, an increased learning rate was found for discriminations along the previously predictive dimension compared with a dimension that was indicative of uncertainty, but again the transfer was more pronounced for specific stimuli that were compatible with the previously learned discrimination. Taken together, the results imply that a transfer of associability typically applies to individual stimuli and depends on how the transfer stimuli relate to those stimuli that individuals previously learned to attend.

Keywords: associative learning, transfer, dimensional associability, attention, contingency learning

Many theories of associative learning are based on the assumption that learning is driven by a competitive attentional mechanism. The associability of a stimulus, and thus the rate of learning, is thought to depend on the allocation of attention to all stimuli that are presented together with a to-be-predicted outcome. Specifically, according to most attentional models of associative learning (e.g., Kruschke, 2001; Mackintosh, 1975; Pearce & Hall, 1980), the degree of learning is expected to depend on the ratio of attention that is paid to a conditioned stimulus and the attention that is paid to other concurrently presented stimuli. The amount of selective attention that is directed to a particular stimulus is assumed to be a function of the associative history of the stimulus.

According to the Pearce and Hall (1980) model, for instance, processing advantages are assumed for stimuli that produced prediction errors over stimuli that allowed accurate outcome predictions. That is, associability will decrease for good predictors of an outcome, whereas uncertainty (e.g., because of partial reinforcement; Kaye & Pearce, 1984) will lead to an increase in associability. Specifically, the change in associative strength of a Stimulus A (ΔV_A, i.e., learning) is determined by Equation 1, and the associability of that stimulus is expected to change on each trial by an amount equal to the absolute difference between the asymptote of conditioning (λ; which depends on properties of the outcome) and the sum of the associative strengths of all stimuli that are present on trial *n* (see Equation 2):

$$\Delta V_A = \theta \alpha_A \lambda \tag{1}$$

$$\alpha_{A,n+1} = \left| \lambda_n - \sum^{n} V_{A,n} \right| \tag{2}$$

The Pearce and Hall model thus assumes the associability parameter to be driven by the absolute prediction error, with increases in associability being expected for stimuli that are low in predictiveness.

In contrast, the Mackintosh (1975) model assumes the associability of Stimulus A to increase if it has been experienced as a more reliable predictor of an outcome compared with other stimuli that were presented, whereas the associability of A decreases if it was a weaker predictor than other stimuli that co-occurred with A. In principle, according to this model (which is most relevant for the present work), the change in associative strength (rather than associability parameter, as suggested by Pearce & Hall, 1980) is assumed to be a function of the prediction error that is made on each trial (see Equation 3 and Bush & Mosteller, 1951):

$$\Delta V_A = \theta \alpha_A (\lambda - V_A) \tag{3}$$

In Equation 3, learning (ΔV_A) depends on the discrepancy between the actual outcome (λ) and the outcome that was predicted by A (i.e., its associative strength V_A), with a small discrepancy $\lambda - V_A$ indicating that A is a reliable predictor. The learning rate parameter \theta is determined by properties of the outcome, and α_A refers to a variable cue processing parameter. In the Mackin-

This article was published Online First November 16, 2015.

Florian Kattner and C. Shawn Green, Department of Psychology, University of Wisconsin–Madison.

Correspondence concerning this article should be addressed to Florian Kattner, Department of Psychology, University of Wisconsin–Madison, 1202 W. Johnson Street, Madison, WI 53706-1611. E-mail: kattner@wisc.edu

tosh (1975) model, α refers to the associability of A (α_A), and it is adjusted on each learning trial according to Equation 4:

$$\begin{aligned} \Delta\alpha_A > 0 \text{ if } |\lambda - V_A| < |\lambda - V_X| \\ \Delta\alpha_A < 0 \text{ if } |\lambda - V_A| \geq |\lambda - V_X| \end{aligned} . \quad (4)$$

Thus, the associability of a Stimulus A increases if A is a better predictor of the outcome than all stimuli that are presented together with A (V_X refers to the sum of the associative strengths of the other stimuli), and the rate of learning is governed by a competitive comparison of the predictiveness of stimuli (relative predictiveness).

There are several human contingency learning studies confirming the assumption that the change in associability of a stimulus depends on its predictiveness relative to other stimuli. A typical example is the *learned predictiveness paradigm*, which assesses the transfer of associability to a new learning stage (Le Pelley & McLaren, 2003). In a first stage, participants learn to respond to compounds of two stimuli, with one stimulus being perfectly predictive of the correct response and the other stimulus being irrelevant. In a second stage, participants have to learn the responses to new compounds consisting of either previously predictive or nonpredictive stimuli (and with both types of compounds now being equally predictive). In line with the Mackintosh (1975) model, it has been shown repeatedly with this paradigm that contingencies with previously predictive stimuli are acquired more readily (indicating the cues' higher associabilities α_A) compared with previously nonpredictive stimuli (e.g., Bonardi, Graham, Hall, & Mitchell, 2005; Griffiths & Le Pelley, 2009; Kattner, 2015; Le Pelley, Beesley, & Griffiths, 2011; Le Pelley & McLaren, 2003; Livesey & McLaren, 2007).

Additional support for the role of associability comes from research on the intradimensional–extradimensional shift effect (see George & Pearce, 1999). Here, stimuli typically vary on two independent dimensions (e.g., color and shape), with one dimension being relevant for the training discriminations, and either the same (intradimensional) or the other dimension (extradimensional) being relevant during transfer (different stimuli are used in training and transfer, thus performance cannot be influenced by direct transfer). In line with Equation 4, several studies reported enhanced learning rates with an intradimensional shift compared with an extradimensional shift (e.g., Roberts, Robbins, & Everitt, 1988; Whitney & White, 1993). This effect could be easily accounted for by assuming that individuals learn to attend to the predictive dimension and to ignore the irrelevant dimension, which would imply that the associability α_A applies to the level of entire stimulus dimensions (as suggested by Sutherland & Mackintosh, 1971). The superiority of an intradimensional shift, however, does not always seem to be true (see Dias, Robbins, & Roberts, 1996; Trobalon, Miguelez, McLaren, & Mackintosh, 2003). According to the Mackintosh (1975) model, learning is assumed to produce changes in the associability of particular stimuli, rather than stimulus dimensions, which might then generalize to new stimuli depending on their similarity. Hence, intradimensional shift effects can potentially be explained in terms of stimulus similarity, as features from the same dimension are expected to be more similar than features from different dimensions.

Trobalon et al. (2003) conducted a spatial discrimination learning study that allowed these two alternative hypotheses to be tested. Rats were trained on either a spatial discrimination task (i.e., arms pointing to specific directions) or a visuotactile discrimination task (with distinctive floor coverings). The transfer task was a spatial discrimination task in which the rats learned to discriminate between arms pointing southeast and southwest. Thus, the transfer discrimination was an intradimensional shift for rats that were trained on the spatial discrimination task, whereas it was an extradimensional shift for rats that were trained on visuotactile discriminations. An enhanced rate of learning following the intradimensional shift was found only if the rats were trained on north versus east or west discriminations (i.e., transfer and training locations did not overlap), whereas south versus east or west discriminations (i.e., the transfer locations lay between the trained locations) resulted in an even slower learning rate for the intradimensional shift compared with the extradimensional shift group (visuotactile training). These results are better explained in terms of stimulus-specific associability (Mackintosh, 1975; see Suret & McLaren, 2003, 2005, for implementations of an algorithm incorporating this principle into an elemental model) than in terms of dimensional associability (Sutherland & Mackintosh, 1971). Specifically, some rats were trained to attend to locations (i.e., east and west) that were similar to the transfer locations (i.e., southeast vs. southwest), whereas other rats had to ignore the transfer locations in order to learn the training discriminations (south vs. east or west). This result is difficult to explain with an account that assumes an increase in associability of the entire "relevant" stimulus dimension (Sutherland & Mackintosh, 1971). The Trobalon et al. (2003) findings are thus in line with the Mackintosh (1975) model, suggesting that associability applies to individual stimuli (e.g., specific locations) rather than to stimulus dimensions in spatial discrimination learning.

There is very little research on human contingency learning focusing on the role of dimensional associability as opposed to stimulus-specific associability. In particular, most of the research on *learned predictiveness* effects has focused on predictiveness varying between individual stimuli rather than between stimulus dimensions. The aim of this study is to evaluate whether associability can apply to stimulus dimensions in human contingency learning. Specifically, the standard learned predictiveness paradigm (Le Pelley & McLaren, 2003) was used to test whether dimensional associability can transfer to a subsequent learning stage. Participants were asked to learn arbitrary responses to perceptually complex Gabor gratings varying on two continuous dimensions (orientation and spatial frequency). To manipulate dimensional predictiveness, only one dimension of the grating was predictive of the response during Stage 1, whereas the other dimension was irrelevant. In Stage 2, participants learned to discriminate two new pairs of Gabor gratings. One pair could be discriminated by attending to the previously predictive dimension, and the other pair could be discriminated by attending to the previously irrelevant dimension. Thus, if dimensional associability transferred to a subsequent learning stage (in line with Sutherland & Mackintosh, 1971), then Stage 2 discriminations that are based on the previously predictive dimension should be acquired more readily.

On the other hand, according to a model that is based on the Pearce and Hall (1980) assumption (see Equation 2), Stage 1 predictiveness might be expected to negatively transfer to a later learning stage. Referring to the Pearce and Hall model, Hall and

Rodriquez (2010) argued that the observed positive transfer effect (learned predictiveness) does not necessarily have to be the result of differences in associability (or learning rate) between predictive and nonpredictive cues. Specifically, these authors argued that training will reduce the associability of predictive stimuli (in line with Pearce & Hall, 1980), but measures of Stage 2 performance in a typical learned-predictiveness experiment may be subject to several factors that can overcome the "negative transfer" effect (e.g., generalization of previously learned associations or biases in the individual's ability to perceptually discriminate stimuli and report these in a later stage). There is indeed some evidence from human contingency learning studies suggesting negative transfer on the level of individual stimuli varying in predictiveness (e.g., Griffiths, Johnson, & Mitchell, 2011, and Hogarth, Dickinson, Austin, Brown, & Duka, 2008, reporting higher learning rates and prolonged eye-gaze times for stimuli associated with uncertain outcomes, respectively). However, negative transfer has never been tested with entire stimulus dimensions varying in predictiveness. As for the present study, an account of dimensional associability that is in line with the Pearce and Hall model would predict higher learning rates at Stage 2 for discriminations that are based on the previously irrelevant dimension.

Here we sought to determine whether the transfer of associability in human contingency learning (whether positive or negative) does apply to entire stimulus dimensions (in line with Sutherland & Mackintosh, 1971). That is, we investigated whether differences in learning rate can be observed for discriminations along stimulus dimensions (i.e., the orientation and spatial frequency of gratings) that had been either relevant or irrelevant for different discriminations in a previous learning stage. A first experiment was designed to replicate the standard learned-predictiveness effect (Le Pelley & McLaren, 2003) at the level of individual stimuli (i.e., higher learning rates for particular gratings that were predictive in a previous learning stage, with both dimensions being equally predictive). Three subsequent experiments then tested whether changes in associability can also be found for entire stimulus dimensions varying in predictiveness (in line with Sutherland & Mackintosh, 1971). Therefore, learning rates were contrasted for discriminations along the previously predictive dimension (intradimensional shift) with discriminations along the previously irrelevant dimension (extradimensional shift).

Experiment 1

Experiment 1 was designed to replicate the basic learned-predictiveness effect (Le Pelley & McLaren, 2003) with a stimulus–response contingency learning paradigm using perceptually complex Gabor-filtered gratings. A replication of the standard transfer effect using the exact same grating stimuli was crucial to assess possible changes in dimensional associability (i.e., the superiority of an intradimensional shift) relative to the stimulus-specific learned-predictiveness effect. The design of Experiment 1 is shown in Table 1. In a first learning stage, participants had to learn particular responses to several compound gratings (e.g., A and X). In line with the Le Pelley and McLaren (2003) design, one grating of each compound was predictive of the required response, whereas the other was not. In a second learning stage, discriminations then had to be learned between new compounds that either contained two previously predictive (e.g., A and C) or two previously nonpredictive (e.g., X and Y) gratings. In line with the Mackintosh (1975) model, faster discrimination learning at Stage 2 is expected for previously predictive gratings because of an increase in associability.

Table 1
Experimental Design of the Two Learning Stages in Experiment 1

Stage 1 Gratings → Response	Stage 2 Gratings → Response
A + X → r1	A + C → r3
A + Y → r1	B + D → r4
B + X → r2	X + V → r3
B + Y → r2	Y + W → r4
C + V → r2	
C + W → r2	
D + V → r1	
D + W → r1	

Note. Each combination of gratings was associated with a particular response (r1 to r4). The capital letters represent individual gratings that were either predictive of the response during Stage 1 (A, B, C, D) or not (V, W, X, Y).

Method

Participants. Twenty-six undergraduate students (20 female, 6 male) were recruited for Experiment 1. Ages ranged between 18 and 25 years (M = 18.7, SD = 1.4). All participants were either compensated with course credit or were paid $10. The entire experiment took about one hour.

Apparatus and stimuli. Stimulus presentation and response registrations were programmed in MATLAB (on a Windows computer) using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Stimuli and text instructions were presented on a 22-in. widescreen TFT monitor with a resolution of 1680 × 1050 pixels. The monitor was placed approximately 59 cm in front of the participant. A standard keyboard was used as the response device.

Eight different monochrome Gabor-filtered gratings (σ = 75, random phase) served as the cues in both learning stages. Each grating was defined by a unique combination of spatial frequency (circles per degree [cpd] of visual angle) and orientation (−60°/2 cpd, −45°/4 cpd, −30°/3 cpd, −15°/5 cpd, 10°/1.5 cpd, 25°/3.5 cpd, 40°/2.5 cpd, and 55°/4.5 cpd). The specific stimulus definitions assured that any two co-occurring gratings always differed in both orientation or spatial frequency. In contrast to previous studies on predictive learning (e.g., using fruit images), however, the use of stimuli that can be manipulated along certain dimensions comes along with a natural ordering of stimuli regarding their similarity (e.g., a 10° and a 25° grating may be more similar than a 10° and a 40° grating). To control for an effect of stimulus similarity, the eight gratings were randomly assigned for each participant to the different types of cues (A, B, C, D, V, W, X, and Y).

Procedure. A typical trial in Experiment 1 is illustrated in Figure 1A. On each trial of Stage 1, two different cues (gratings) were presented side by side on the screen (right and left positions were counterbalanced across trials). The participants were required to press the right or left arrow key, with one key being the correct response and the other key being the incorrect response. Having

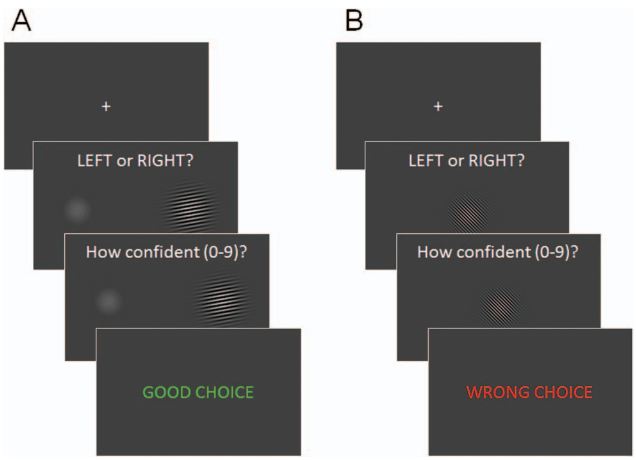


Figure 1. Examples of the trial sequence (A) with two stimuli differing in predictiveness (Experiment 1), and (B) with the two dimensions of a presented stimulus (i.e., orientation and spatial frequency) differing in predictiveness (Experiments 2 to 4). See the online article for the color version of this figure.

given a response, the participants were asked to rate how confident they were that their response was correct by pressing a number of the numeric pad (with 0 being *totally uncertain* and 9 being *totally confident*). Feedback was then presented for 750 ms, indicating to the participant whether the response was correct ("good choice") or incorrect ("wrong choice"). The next trial started after an intertrial interval of 750 ms. Moreover, a current score was presented in the center of the screen during the feedback and the intertrial interval. On each trial the score increased by the amount of the confidence rating if the response was correct, and decreased by the amount of the confidence rating if the response was incorrect. The participants were instructed that their task was to learn the relationship between the cues and the required response in order to increase their score as much as possible (starting with a score of 0).

The individual gratings were assigned to eight different compounds, each of which was associated with a particular response (r1 or r2; the arrow keys were assigned randomly for each participant; see Table 1). Each of the eight combination of cues was repeated 40 times (in blocks), resulting in a total number of 320 trials in Stage 1. The order of the eight types of trials was randomized within each block.

The basic procedure of Stage 2 was identical to Stage 1. However, new compounds consisting of the same individual gratings were presented, and the participants were now required to respond with the up- or down-arrow keys (which were assigned to either response r3 or r4; see Table 1). The score continued from the final score in Stage 1. Each of the four combinations of cues was presented 30 times during Stage 2, resulting in a total of 120 trials. The order of trials was randomized for each repetition.

Results and Discussion

An individual accuracy of more than 50% correct responses in the last block of Stage 1 (five trials per compound) was defined as the minimum criterion for discrimination learning. The data of one participant, who did not meet this lax criterion, were not included in the analysis (although note that the exclusion of data did not affect the overall pattern or the significance of the results).

The remaining 25 participants demonstrated successful learning during Stage 1. An 8 (compound) $\times$ 8 (block) repeated-measures ANOVA on the accuracy of responses confirmed a significant learning effect across blocks, $F(7, 168) = 8.12$, $p < .001$, $\eta^2_G = 0.02$, with an increase from 55.2% ($\pm$11.1%) correct responses in the first block to 71.8% ($\pm$10.1%) correct responses in the eighth block (see Figure 2). There were no significant differences in accuracy between the eight compounds, $F(7, 168) = 0.56$, $p = .79$, as well as no Block $\times$ Compound interaction, $F(49, 1176) = 1.10$, $p = .29$. Likewise, confidence increased significantly from an average rating of 4.82 ($\pm$0.69) in Block 1 to 6.65 ($\pm$0.71) in Block 8, $F(7, 168) = 5.72$, $p < .001$, $\eta^2_G = 0.10$ (see vertical bars in Figure 2), but it did not differ between compounds, $F(7, 168) = 0.49$, $p = .84$, and there was also no interaction, $F(49, 1176) = 0.94$, $p = .58$. Thus, the increases in both accuracy and confidence during Stage 1 did not seem to differ between compounds.

At Stage 2, the overall accuracy increased from 63.0% ($\pm$10.8%) in Block 1 to 79.4% ($\pm$9.1%) in Block 6. A 6 (block) $\times$ 2 (predictiveness at Stage 1) repeated-measures ANOVA confirmed the overall learning effects with a significant main effect of block, $F(5, 120) = 7.51$, $p < .001$, $\eta^2_G = 0.11$. More importantly, there was also a significant main effect of Stage 1 predictiveness on the accuracy of responses in Stage 2, $F(1, 24) = 7.68$, $p = .011$, $\eta^2_G = 0.08$, indicating that the responses to new compounds that consisted of previously predictive gratings were more accurate ($M = 78.0\% \pm 9.3\%$) than the responses to previously nonpredictive gratings ($M = 68.9\% \pm 10.4\%$). In addition, there was a significant Block $\times$ Predictiveness interaction on the accuracy of responses in Stage 2, $F(5, 120) = 3.94$, $p = .002$, $\eta^2_G = 0.05$, suggesting that the learning rate was greater for compounds containing previously predictive gratings than for compounds of previously nonpredictive gratings (see Figure 2). An additional Bayesian analysis (see Morey & Rouder, 2011, and Rouder, Speckman, Sun, Morey, & Iverson, 2009, for the calculation of a Bayes factor) of this learned-predictiveness effect suggested that the alternative hypothesis (i.e., higher accuracy for previously predictive stimuli) was 4.55 times more likely than the null hypothesis.

Confidence ratings at Stage 2 increased from 5.28 ($\pm$0.78) in Block 1 to 7.50 ($\pm$0.54) in Block 6. The confidence ratings were subject to a main effect of block, $F(5, 120) = 12.63$, $p < .001$, $\eta^2_G = 0.29$, but there was no main effect of predictiveness, $F(1, 24) = 0.04$, $p = .84$, and the interaction did not reach significance, $F(5, 120) = 2.10$, $p = .07$, indicating that learned predictiveness had less of an effect on the confidence of responses than on accuracy.

Learning rates were estimated for each participant by fitting a power function, $y = \alpha * x^\beta$ (in line with classical approaches of fitting learning curves; cf. Newell & Rosenbloom, 1981), to the accuracy of discriminations during Stage 2 that were based on either previously predictive or nonpredictive stimuli. The average slope of the learning curve was significantly greater for the accuracy of responses to A + C and B + D ($M_\beta = 0.20$, $SE_\beta = 0.05$) than for responses to X + V and Y + W ($M_\beta = 0.02$, $SE_\beta = 0.05$), $t(23) = 2.49$, $p = .020$ (paired t test), indicating again that

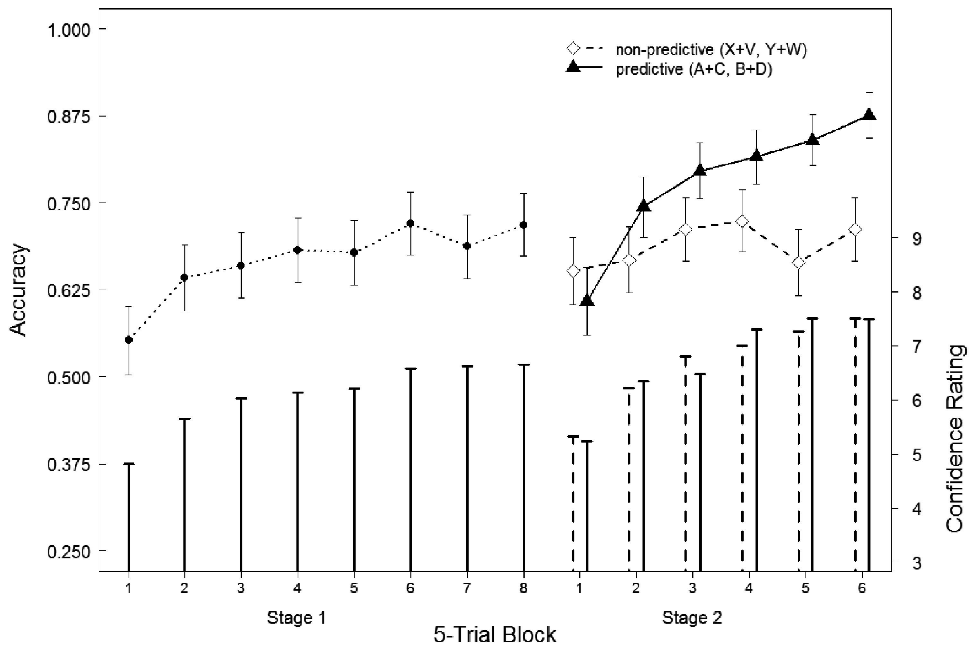


Figure 2. Mean accuracy (left ordinate; points) and confidence ratings (right ordinate; vertical bars) for the responses across five-trial blocks of Stage 1 (left/right) and Stage 2 (up/down) in Experiment 1. Stage 2 data are shown separately for discriminations of compounds consisting of previously predictive (solid lines) and nonpredictive (dashed lines) gratings. Error brackets refer to standard errors of the mean.

discriminations that were based on previously predictive stimuli were acquired more rapidly than those with nonpredictive stimuli. A Bayesian analysis revealed that the alternative hypothesis was 4.45 times more likely than a null difference in slopes.

In line with the Mackintosh (1975) model, Experiment 1 replicated a positive transfer of predictiveness in a stimulus–response contingency learning paradigm using gratings. That is, cue–response associations were learned faster for cues that were predictive of a response in a previous learning stage compared with cues that were nonpredictive of a response. Given this result, the purpose of the following experiments was to assess whether changes in the associability of an entire dimension of Gabor gratings (i.e., either orientation or spatial frequency) occur as a function of its predictiveness.

Experiment 2

In Experiment 2, we tested whether the same type of learned-predictiveness effect – as observed for individual stimuli in Experiment 1 – could be found with entire stimulus dimensions (i.e., either orientation or frequency of the gratings). Therefore, in Experiment 2, only one grating was presented on each trial, with one dimension being indicative of the required response and the other dimension being irrelevant. To investigate changes in dimensional associability, four different gratings were presented in each learning stage. The Stage 2 stimuli could be discriminated either along the previously predictive (intradimensional shift) or nonpredictive (extradimensional shift) stimulus dimension. The design of Experiment 2 is shown in Table 2.

Method

Participants. Twenty-five undergraduate students (13 female, 12 male) participated in Experiment 2 for course credit or payments of $10. Ages ranged between 18 and 27 years ($M = 21.2$, $SD = 2.4$). The experiment took about one hour.

Apparatus and stimuli. The same apparatus was used as in Experiment 1.

Five different orientations (−45°, −30°, 5°, 15°, 30°) and spatial frequencies (1, 2, 3, 4, and 5 cpd) were used to create the eight types of Gabor gratings ($\sigma = 75$, random phase). The base color of the gratings was varied by choosing RGB values from a uniform random distribution on each trial of both learning stages. Thus, color was an entirely unpredictive dimension in Experiment 2.

Table 2
Experimental Design of the Learning Stages in Experiments 2, 3, and 4

Stage 1 Grating → Response	Stage 2 Grating → Response
a1 → r1	c3 → r3
a2 → r1	d3 → r4
b1 → r2	e4 → r3
b2 → r2	e5 → r4

Note. Each individual grating was associated with a particular response (r1 to r4). A grating is defined by letter and a number. Letters represent a value along the stimulus dimension that was predictive of the required response during Stage 1, and numbers represent the dimension that was irrelevant during Stage 1 (see text for further details).

The eight different types of gratings that were presented in Stages 1 and 2 of Experiments 2 to 4 are depicted in Table 2. Each grating is represented as a combination of a lowercase letter and a digit, with the letter (a through e) referring to a value along the dimension that was predictive in Stage 1, and the digit (1 through 5) referring to values along the nonpredictive dimension. Note that neither the alphabetical position of the letter nor the digit refers to the position along a particular dimension.

Procedure. The procedure was very similar to Experiment 1, except that only a single grating was presented on each trial, with one of two dimensions (either spatial frequency or orientation) being predictive of the correct response. Again, the participants' task was to press either the left- or the right-arrow key, with a particular response being mapped to each grating (see Table 2). Participants could thus in principle learn the correct stimulus associations by attending to entire stimulus dimensions. Figure 1B depicts the sequence of a typical trial in Experiment 2 (which was identical in Experiments 3 and 4).

In Stage 1, the same four gratings were presented to all participants (i.e., all four possible combinations of −45°, 30°, 2 cpd, and 4 cpd). These gratings were assigned randomly to the four types of cues (a1, a2, b1, and b2) in a way that one dimension was predictive of the response, whereas the other dimension was unpredictive of the response. Hence, for 11 participants, spatial frequency was predictive (e.g., a1 = 30°/2 cpd; b1 = 30°/4 cpd), and for 14 participants, orientation was predictive during Stage 1 (e.g., a1 = −45°/2 cpd; b1 = 30°/2 cpd). Each type of trial was repeated 40 times during Stage 1, resulting in a total of 160 trials.

Stage 2 contained four new gratings as defined by different orientations and frequencies (i.e., −30°, 5° or 15°, and 1, 3, or 5 cpd). Random combinations of these orientations and spatial frequencies were assigned to the different types of cues in a way that two gratings could be discriminated on the basis of the previously predictive dimension (c3 vs. d3), whereas the other two gratings could be discriminated on the basis of the previously nonpredictive dimension (e4 vs. e5). The specific gratings that were presented to each participant in the two learning stages are illustrated in Appendix A. In order to avoid ceiling effects (as a result of the low number of stimuli), the stimulus–response associations at Stage 2 were slightly probabilistic, $P(\text{response}|\text{cue}) = .9$. That is, each cue was associated with the respective response in 90% of the trials, and with the opposite response in 10% of the trials. Each type of trial was repeated 30 times, resulting in a total of 120 trials in Stage 2.

Results and Discussion

Two participants (both with spatial frequency being relevant at Stage 1) did not reach the learning criterion for Stage 1 (accuracy in Block 8 > 50%), and the data analysis was based on the remaining 23 participants (the pattern of results did not depend on whether the nonlearners were included or not).

Significant learning was observed during Stage 1, with 73.9% (±9.2%) correct responses in Block 1, and 93.0 (±5.3%) correct in Block 8 (see Figure 3). The confidence of responses increased from 5.46 (±0.72%) in Block 1 to 8.15 (±0.48%) in Block 8. Separate 8 (block) × 4 (grating) repeated-measures ANOVAs confirmed this learning effect across blocks for both accuracy, $F(7, 154) = 8.78$, $p < .001$, $\eta^2_G = 0.12$, and confidence ratings, $F(7, 154) = 14.00$, $p < .001$, $\eta^2_G = 0.32$. There was no main effect of

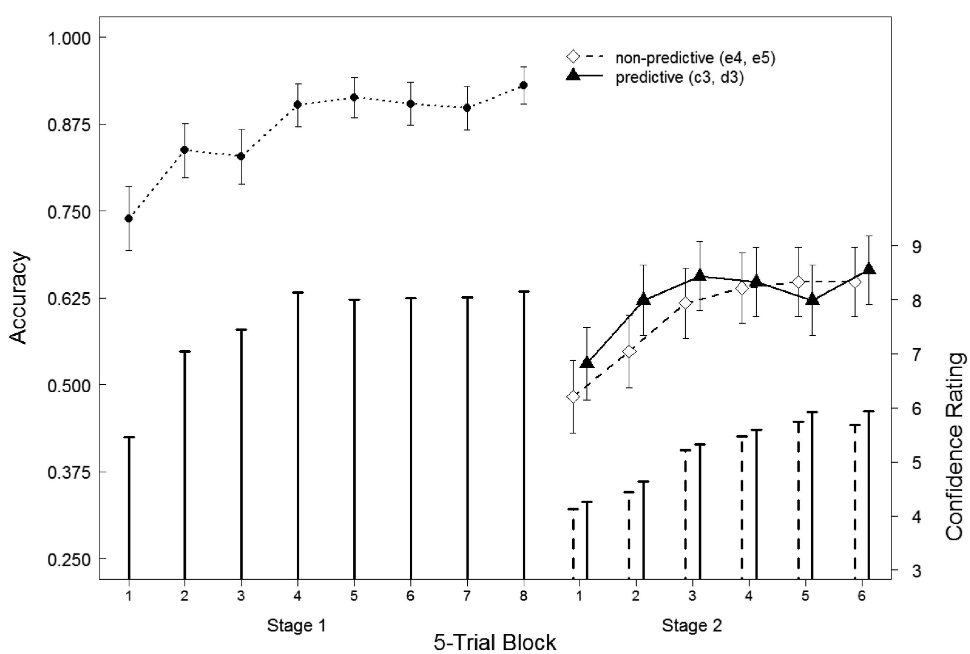


Figure 3. Mean accuracy (left ordinate; points) and confidence ratings (right ordinate; vertical bars) for the responses across five-trial blocks of Stage 1 (left/right) and Stage 2 (up/down) in Experiment 2. Stage 2 data are shown separately for discriminations based on the previously predictive (solid lines) and nonpredictive (dashed lines) stimulus dimension. Error brackets represent standard errors of the mean.

grating on either accuracy, $F(3, 66) = 0.23$, $p = .88$, or confidence during Stage 1, $F(3, 66) = 1.00$, $p = .40$. The Block × Grating interactions were also not significant for accuracy, $F(21, 462) = 0.94$, $p = .54$, and confidence, $F(21, 462) = 1.14$, $p = .30$, respectively. Thus, learning rates at Stage 1 did not appear to differ between the four gratings.

The average accuracy of responses in Stage 2 increased from 50.7% (±10.4%) correct in Block 1 to 65.7% (±9.9%) in Block 6. Likewise, confidence ratings increased from 4.20 (±0.75%) in Block 1 to 5.80 (±0.73%) in Block 6 (see Figure 3). Two 6 (block) × 2 (predictiveness in Stage 1) repeated-measures ANOVAs confirmed these learning effects with significant main effects of block for both accuracy, $F(5, 110) = 5.15$, $p < .001$, $\eta^2_G = .08$, and confidence, $F(5, 110) = 4.17$, $p = .002$, $\eta^2_G = 0.12$. Importantly, however, there was no main effect of Stage 1 predictiveness on either accuracy, $F(1, 22) = 0.37$, $p = .55$, or confidence $F(1, 22) = 0.28$, $p = .60$, indicating that the previous predictiveness of an entire stimulus dimension did not transfer to Stage 2. There were also no Block × Predictiveness interactions, $F(5, 110) = 0.54$, $p = .75$ (accuracy), and $F(5, 110) = 0.04$, $p > .99$ (confidence), suggesting that the rate of learning did not depend on the previous predictiveness of the discriminative dimension. An additional Bayesian analysis (Morey & Rouder, 2011; Rouder et al., 2009) confirmed that the null hypothesis (i.e., no difference in accuracy as a function of dimensional predictiveness) was 3.87 times more likely (Bayes factor scaled to Jeffrey-Zellner-Siow prior, $r = .707$) than the alternative hypothesis.

Curve fitting of individual data provided separate learning rate estimates for Stage 2 discriminations that were based on the previously predictive (c3 vs. d3) and nonpredictive (e4 vs. e5) stimulus dimension. In contrast to Experiment 1, the slopes of the learning curve did not differ significantly between c3 vs. d3 ($M_\beta = 0.13$, $SE_\beta = 0.06$) and e4 vs. e5 discriminations ($M_\beta = 0.17$, $SE_\beta = 0.05$), $t(22) = -0.50$, $p = .62$ (paired t test). The Bayes factor indicated that the null hypothesis (i.e., equal slopes) was 4.08 times more likely than a difference in learning rates.

The results of Experiment 2 suggest that stimulus dimensions (orientation and frequency) may not be subject to the same type of learned-predictiveness effect that has been reported with individual stimuli (e.g., Le Pelley et al., 2011; Le Pelley & McLaren, 2003). In particular, the learning rate of Gabor discriminations at Stage 2 did not depend on whether the discriminatory dimension had been predictive at Stage 1. The absence of transfer of predictiveness for stimulus dimensions indicates that the acquired associability at Stage 1 may not be linked to the (predictive) dimensions of a cue, but rather to the entire configuration of stimulus dimensions (i.e., the individual stimulus). That is, in contrast to the suggestions made by Sutherland and Mackintosh (1971), Experiment 2 does not seem to provide any indication for associability to apply to stimulus dimensions in human contingency learning.

However, Trobalon et al. (2003) demonstrated that the type of transfer observed depends on the similarity between the stimuli that are used in Stage 1 and Stage 2. In particular, they found no evidence for positive transfer of predictiveness when the to-be-discriminated locations at Stage 2 lay between those that were discriminated in Stage 1, suggesting that the associability of the test locations was lower (i.e., the stimuli were harder to discriminate) compared with the trained locations. In contrast, with orthogonal locations, Trobalon et al. observed positive transfer in line with the Mackintosh (1975) predictions. It may thus be crucial to look at the similarity relations between the two learning stages of the present experiment. As a consequence of the random assignment of orientations and spatial frequencies (which was done to control for effects of stimulus order), the values that had to be discriminated at Stage 2 (i.e., c and d) could, in different participants, be either more or less extreme values along the dimension that was predictive in Stage 1. According to an account of stimulus-specific associability (Mackintosh, 1975), the observed absence of a learned-predictiveness effect might be related to the low associability of stimulus values that participants learned to ignore during Stage 1 (i.e., orientations or frequencies that lay between those that were discriminated).

With the specific orientations that were chosen for Stage 1 (−45° and 30°), the to-be-discriminated orientations at Stage 2 were always in between the trained orientations ($n = 14$). Likewise, the majority of participants ($n = 7$) who learned spatial-frequency discriminations at Stage 1 were not confronted with more extreme frequency discriminations at Stage 2 (e.g., c = 3 cpd vs. $d = 5$ cpd). These 21 participants might thus have learned to ignore the intermediate values that would be relevant in Stage 2, while having attended to the more extreme values that were relevant in Stage 1. Only two participants learned more extreme frequency discriminations at Stage 2 (c = 1 cpd vs. $d = 5$ cpd; compared with a = 2 cpd vs. b = 4 cpd at Stage 1). According to Mackintosh (1975), the associability of the spatial frequencies that were relevant in Stage 2 should thus have been high. As an additional between-subjects factor, the type of stimulus similarity (intermediate orientations, intermediate frequencies, more extreme frequencies) had no main effect on accuracy, $F(2, 20) = 1.07$, $p = .36$, nor did it interact with predictiveness, $F(2, 20) = 0.11$, $p = .90$, or block, $F(10, 100) = 0.51$, $p = .88$. There was also no Similarity × Predictiveness × Block interaction, $F(10, 100) = 0.97$, $p = .47$. However, because of the low number of participants with more extreme values at Stage 2 ($n = 2$), this null effect does not rule out the possibility that the transfer of predictiveness may depend on the similarity relations.

The observed absence of a transfer effect for dimensional associability may thus be in line with Trobalon et al. (2003), suggesting that transfer of associability depends on the specific relations between the stimuli that are relevant during training and during test. For most of the participants in Experiment 2, the associability of the Stage 2 stimuli may have been low because they were to be ignored during Stage 1. This finding clearly argues against the assumption that participants learn to attend to the entire "relevant" dimension.

Experiment 3

Experiment 3 was an attempt to replicate the results of Experiment 2 with a slightly larger sample and some minor modifications. Specifically, participants were presented with probabilistic cue–response associations at Stage 1 ($p = .85$), which might maintain a certain level of uncertainty (participants in Experiment 2 might have started to neglect the predictive dimension because of the absence of prediction errors; cf. Pearce & Hall, 1980). Thus, Experiment 3 tested whether the absence of a transfer of dimensional associability can be found with slightly elevated attention on the relevant stimulus dimension.

Method

Participants. Another sample of 34 undergraduate students (19 female, 15 male) participated in Experiment 3. Ages ranged between 18 and 28 years (M = 20.1, SD = 2.2). All participants were compensated with course credit or payments of $10. The experimental session took about one hour.

Materials and procedure. Stimuli, apparatus, and procedure were identical to Experiment 2 (see Table 1), with the only exception being that the cue–response associations were probabilistic at Stage 1, $P(\text{response}|\text{cue}) = .85$. That is, a particular grating (e.g., a1) was associated with the respective response (r1) in 85% of the cases (34 trials), whereas it was associated with the opposite response (r2) in 15% of the cases (six trials). In Stage 2, each cue was associated with the respective response in 90% of the trials, and with the opposite response in 10% of the trials, $P(\text{response}|\text{cue}) = .9$. Again, each type of trial was repeated 40 times during Stage 1, and 30 times during Stage 2, resulting in a total of 160 and 120 trials, respectively.

The gratings were assigned to the different types of cues according to the same random procedure as in Experiment 2. The same four gratings were presented to all participants in Stage 1 (i.e., −45° or 30° × 2 or 4 cpd), with one of the two dimensions being predictive of the response, whereas random combinations of the remaining values of orientations and spatial frequencies were presented in Stage 2. Spatial frequency was the predictive dimension in Stage 1 for 20 participants, and orientation was predictive for 14 participants. The specific gratings that were presented to each participant in Experiment 3 are illustrated in Appendix B.

Results and Discussion

Five participants in Experiment 3 (four with Stage 1 discriminations based on spatial frequency) did not reach an accuracy greater than 50% in the final block of Stage 1. The following analysis is based on the data from the remaining 29 participants (the significance of results did not depend on the exclusion of participants).

During Stage 1, the accuracy of discriminations increased from 57.2% (±9.2%) correct in Block 1 to 73.8% (±8.2%) correct in Block 8 (see Figure 4). The average confidence ratings grew from 3.63 (±0.57) in Block 1 to 6.31 (±0.60) in Block 8. Because of the probabilistic nature of the stimulus–response associations, the participants' confidence of the responses throughout Stage 1 was significantly lower compared with Experiment 2, $t(57) = 13.53$, $p < .001$, $\eta^2_G = 0.19$ ($M_{\text{Exp.2}} = 7.2$, $M_{\text{Exp.3}} = 5.2$). Learning effects were again confirmed by main effects of block for accuracy, $F(7, 196) = 5.93$, $p < .001$, $\eta^2_G = 0.05$, and confidence, $F(7, 196) = 10.51$, $p < .001$, $\eta^2_G = 0.20$. There were no main effects of grating, $F(3, 84) = 0.62$, $p = .60$ (accuracy), and $F(3, 84) = 0.28$, $p = .84$ (confidence). The Block × Grating interactions were not significant for either measure, $F(21, 588) = 1.56$, $p = .053$ (accuracy), and $F(21, 588) = 1.37$, $p = .12$ (confidence). Thus, the learning rates in Stage 1 did again not differ between the four gratings.

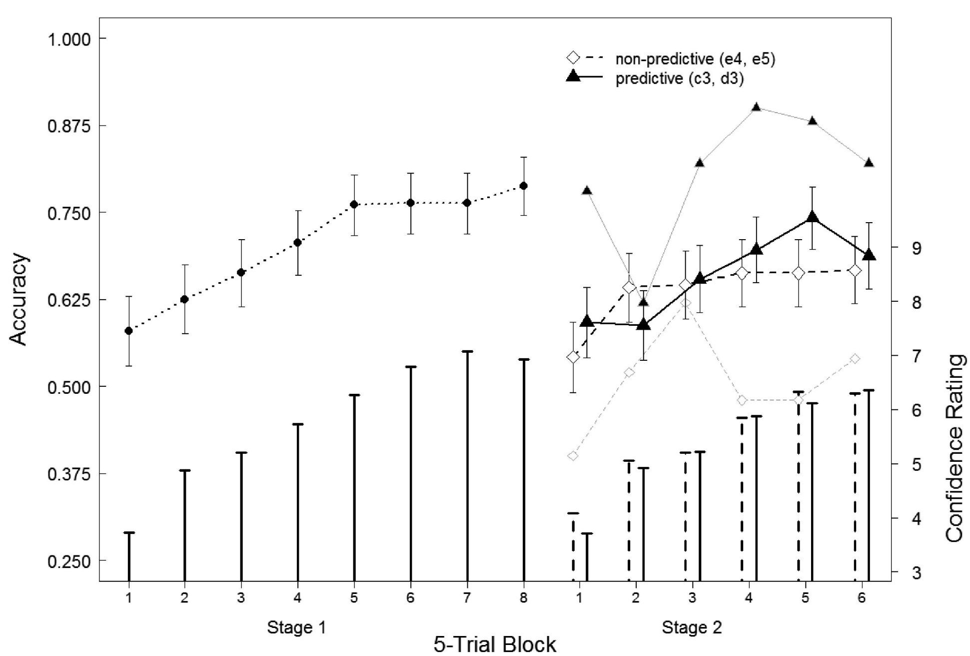


Figure 4. Mean accuracy (left ordinate; points) and confidence ratings (right ordinate; vertical bars) for the responses across five-trial blocks of Stage 1 (left/right) and Stage 2 (up/down) in Experiment 3. Stage 2 data are shown separately for discriminations based on the previously predictive (solid lines) and nonpredictive (dashed lines) stimulus dimension. Error brackets indicate standard errors of the mean. The thin lines depict the average accuracy for participants with decreased similarity along the predictive dimension (see the Results and Discussion sections).

In Stage 2, the average accuracy of responses increased from 55.9% (±9.2%) correct in Block 1 to 65.0% (±8.9%) correct in Block 6, with the confidence ratings simultaneously increasing from 3.61 (±0.58) to 5.94 (±0.57). A 6 (block) × 2 (predictiveness) repeated-measures ANOVA confirmed a significant main effect of block for accuracy, $F(5, 140) = 4.02$, $p = .002$, $\eta^2_G = 0.05$, as well as for the confidence ratings, $F(5, 140) = 13.06$, $p < .001$, $\eta^2_G = 0.25$. However, the accuracy of responses in Stage 2 was again not subject to a main effect of prior predictiveness, $F(1, 28) = 0.22$, $p = .64$, and there was also no interaction, $F(5, 140) = 1.26$, $p = .29$, indicating that there were no increases in dimensional associability with a sustained level of uncertainty (see Figure 4). Likewise, there was no main effect of predictiveness, $F(1, 28) < .01$, $p = .99$, and no interaction, $F(5, 115) = 6.35$, $p = .67$, on the confidence ratings. The Bayes factor suggests that the null hypothesis was 4.57 times more likely than the alternative hypothesis of Stage 1 predictiveness to influence the accuracy of responses at Stage 2 (Jeffrey-Zellner-Siow prior, $r = .707$).

Likewise, the learning rate estimates obtained from power functions fitted to individual accuracy at Stage 2 did not differ significantly between discriminations that were based on the previously predictive ($M_\beta = 0.11 \pm 0.04$) and nonpredictive dimension ($M_\beta = 0.08 \pm 0.04$), $p = .50$ (paired t test); Bayes factor in favor of the null = 4.07.

Experiment 3 replicated the results of Experiment 2 with probabilistic cue–response associations at Stage 1. This indicates that even with lower levels of maximum performance at Stage 1 (comparable with those in Experiment 1), we did not find transfer effects for the predictiveness of stimulus dimensions in Experiment 3. The reduction of uncertainty to a minimum at the end of Stage 1 was thus most likely not the cause of the absence of a learned-predictiveness effect in Experiment 2. In other words, we did not find any evidence of transfer of dimensional associability in human contingency learning with deterministic and probabilistic stimulus–outcome associations.

On the other hand, according to Mackintosh (1975), the transfer of stimulus-specific associability may again depend on the relations between the exact stimuli presented in Stages 1 and 2. With the type of similarity (intermediate orientations, intermediate frequencies, more extreme frequencies) as an additional between-subjects factor, the ANOVA on Stage 2 accuracy revealed a significant interaction between similarity and prior predictiveness, $F(2, 26) = 14.58$, $p < .001$, $\eta^2_G = 0.12$, indicating that the learned-predictiveness effect was qualified by the specific stimuli that were presented at Stage 2.

Specifically, with the random assignments of stimulus values, five participants in Experiment 3 learned Stage 2 discriminations of spatial frequencies that were more extreme than those that were predictive in Stage 1. The Mackintosh (1975) model predicts these stimuli to be high in associability, because they would have been highly predictive of the discrimination in Stage 1. The average Stage 2 performance of these five participants is illustrated as thin gray lines in Figure 4. A separate 6 (block) × 2 (predictiveness) repeated-measures ANOVA revealed that accuracy differed significantly between discriminations along the previously predictive ($M = 0.80$) and nonpredictive dimension ($M = 0.51$), $F(1, 4) = 15.50$, $p = .017$, $\eta^2_G = 0.48$, indicating a positive transfer of predictiveness for participants with more extreme spatial-frequency discriminations at Stage 2 (i.e., 1 vs. 5 cpd).[1] There was no significant main effect of block, $F(5, 20) = 1.49$, $p = .24$, and no interaction, $F(1, 4) = 1.31$, $p = .30$. For the remaining 11 participants with frequency being the predictive dimension, one frequency value lay between the predictive values that were presented in Stage 1. Likewise, when orientation was predictive at Stage 1 (13 participants), the Stage 2 orientations lay between the Stage 1 values. In both cases, a stimulus-specific account predicts low associability for the specific orientations and frequencies that were presented at Stage 2. Accordingly, the ANOVA on accuracy revealed only a significant main effect of block, $F(5, 115) = 3.48$, $p = .006$, $\eta^2_G = 0.06$, but no main effect of prior predictiveness, $F(1, 18) = 1.85$, $p = .19$, and no interaction, $F(5, 90) = 0.71$, $p = .62$.

Thus, Experiment 3 replicated and extended the findings of Experiment 2. A transfer of dimensional associability was absent for the majority of participants for whom the to-be-ignored stimuli at Stage 1 lay between those that were relevant at Stage 2. By contrast, those few participants that were presented with more extreme instantiations along the previously predictive dimension showed a positive transfer of associability with regard to the previously predictive dimension. This observation is in line with previous animal-learning data (Trobalon et al., 2003), and it suggests that associability is not necessarily enhanced for an entire dimension that is relevant for a discrimination problem. Rather, associability in human contingency learning seems to change for particular stimuli and transfer or generalize as a function of the similarity relations.

According to the Pearce and Hall (1980) model, associability should increase with the degree of uncertainty associated with a stimulus, whereas stimuli that perfectly predict an outcome should lose associability. Although this account seems to be incompatible with stimulus-specific learned-predictiveness effects (as found in Experiments 1 to 3), there is also evidence that uncertainty sometimes does enhance associability in human contingency learning (e.g., Griffiths et al., 2011; Hogarth et al., 2008). It has been argued that associability may in fact be based on separate predictiveness-driven and uncertainty-driven mechanisms (e.g., Esber & Haselgrove, 2011; George & Pearce, 2012; Le Pelley, 2004; Pearce & Mackintosh, 2010). In particular, the two mechanisms might apply to different stages or levels of stimulus processing. For instance, George and Pearce (2012) argued that the predictiveness-driven mechanism should affect the salience of individual stimuli, whereas the uncertainty-driven mechanism drives the associability of stimulus configurations. There might also be differences with regard to the mechanisms that apply to stimulus-specific and dimensional associability.

Experiment 4 was an attempt to find evidence for an uncertainty-based learning mechanism influencing dimensional associability. In particular, we tested whether transfer was observed for uncertainty information that was provided by a particular stimulus dimension. Experiment 3 showed that uncertainty of the stimulus–response associations does not lead to a positive transfer

[1] Note that there was only one participant (No. 23; see Appendix B) for which spatial frequency was irrelevant in Stage 1 (i.e., orientation being relevant), and 1 versus 5 cpd discriminations in Stage 2. This participant reached a mean accuracy of 0.82 in Stage 2.

of dimensional predictiveness. However, if the uncertainty-driven learning mechanism applied to the level of dimensional associability, then the uncertainty information provided by a stimulus dimension might transfer to a subsequent learning stage.

Experiment 4

The aim of Experiment 4 was to assess the transfer of uncertainty information to a subsequent learning stage. Stimulus–response associations were again assigned such that one stimulus dimension was predictive of the response, whereas the other dimension was irrelevant for the decision of which response to choose. However, the "irrelevant" dimension provided information on whether the stimulus–response association was deterministic or probabilistic. That is, the "irrelevant" dimension indicated the level of uncertainty that was associated with the required response (without being indicative of the response itself). If an uncertainty-driven learning mechanism (Pearce & Hall, 1980) applied to the level of entire stimulus dimensions, then changes in associability might be observed as a function of whether the dimension provides uncertainty information.

Method

Participants. Thirty-eight undergraduate students (23 female, 15 male) participated in Experiment 4. Ages ranged between 18 and 48 years ($M = 22.7$, $SD = 6.1$). The entire experiment took about one hour, and participants were either compensated with course credit or they received a payment of $10.

Materials and procedure. Experiment 4 utilized the same apparatus and stimuli as Experiment 3.

The basic experimental design was identical to Experiments 2 and 3 (see Table 1), with a1 and a2 being predictive of r1, and b1 and b2 being predictive of r2. However, two cues (a1 and b1) perfectly predicted the respective response, $P(r1|a1) = P(r2|b1) = 1$, whereas the other two cues (a2 and b2) were probabilistically associated with the response, $P(r1|a2) = P(r2|b2) = .7$. That is, one stimulus dimension (i.e., the dimension that was not predictive of a particular response) was indicative of the uncertainty associated with the responses given during Stage 1. According to the Pearce and Hall (1980) assumption, the information regarding uncertainty might affect the associability of the respective stimulus dimension, and hence lead to differential learning rates at Stage 2. As in Experiments 2 and 3, the cue–response associations during Stage 2 were also probabilistic, $P(\text{response}|\text{cue}) = .9$.

The specific gratings were again randomly assigned to the different types of cues in the same way as in Experiments 2 and 3. That is, the same four gratings were presented to all participants in Stage 1 (but as different types), and random combinations of the remaining orientations and frequencies were presented in Stage 2. Consequently, 21 participants learned spatial-frequency discriminations and 17 participants learned orientation discriminations during Stage 1. The particular orientations and spatial frequencies of the gratings that were presented to each individual in the two learning stages are illustrated in Appendix C.

Results and Discussion

Seven participants in Experiment 4 did not reach the 50% criterion for successful learning in the final block of Stage 1, and their data were not included in the analysis (the significance of any effects did not depend on subject exclusions). For the remaining 31 participants, the accuracy of the responses in Stage 1 increased from 59.4% (±8.8%) correct on average in Block 1 to 73.1% (±8.0%) in Block 8. Likewise, confidence grew from 3.24 (±0.53) in Block 1 to 6.46 (±0.61) in Block 8 (see Figure 5). Separate 8 (block) × 2 (probability of outcome: $p = 1$, $p = .7$) repeated-measures ANOVAs confirmed significant main effects of block on accuracy, $F(7, 210) = 4.99$, $p < .001$, $\eta_G^2 = 0.07$, and confidence, $F(7, 210) = 21.30$, $p < .001$, $\eta_G^2 = 0.36$. Moreover, there were significant main effects of the probability of the outcome on accuracy, $F(1, 30) = 45.25$, $p < .001$, $\eta_G^2 = 0.23$, as well as on confidence, $F(1, 30) = 7.44$, $p = .011$, $\eta_G^2 = 0.01$, indicating that, as expected, accuracy and confidence was higher for deterministic associations (78.7% correct; mean confidence rating = 5.61) than for probabilistic associations (60.9% correct; mean confidence rating = 5.25). There was a significant interaction on accuracy, $F(7, 210) = 2.81$, $p = .008$, $\eta_G^2 = 0.03$ (indicating a higher learning rate for deterministic than for probabilistic associations), but not on confidence, $F(7, 210) = 0.67$, $p = .70$.

In Stage 2, the average accuracy increased from 61.3% (±8.8%) correct responses in Block 1 to 64.8% (±8.6%) in Block 6. Average confidence ratings increased from initial 4.03 (±0.61) to 6.30 (±0.56). A 6 (block) × 2 (predictiveness in Stage 1) repeated-measures ANOVA on accuracy revealed no significant main effect of block, $F(5, 150) = 0.99$, $p = .43$, but a significant Predictiveness × Block interaction, $F(5, 150) = 6.53$, $p < .001$, $\eta_G^2 = 0.06$, indicating that learning rate differed between discriminations that were based on the previously predictive (c3 vs. d3) and irrelevant stimulus dimension (e4 vs. e5). Particularly, the accuracy was higher for c3 vs. d3 discriminations than for e4 vs. e5 discriminations in the first four blocks ($.003 < p < .03$), whereas there were no significant differences in accuracy in Block 5 ($p = .50$) and Block 6 ($p = .08$). In addition, there was also a main effect of prior c3 vs. d3 predictiveness on accuracy, $F(1, 30) = 5.02$, $p = .033$, $\eta_G^2 = 0.05$, suggesting that accuracy was higher for discriminations that were based on the previously predictive dimension ($M = 65.4 \pm 8.5\%$) than for discriminations based on the previously irrelevant (and uncertainty-indicating) dimension ($M = 58.0 \pm 8.9\%$). The calculation of a Bayes factor (Jeffrey-Zellner-Siow prior, $r = .707$) confirmed that the alternative hypothesis of a learned-predictiveness effect on accuracy was 4.29 times more likely than the null hypothesis. Confidence ratings at Stage 2 were subject to a main effect of block, $F(5, 150) = 12.44$, $p < .001$, $\eta_G^2 = 0.23$, but there was no main effect of predictiveness, $F(1, 30) = 1.60$, $p = .22$, and no interaction, $F(5, 150) = 1.48$, $p = .20$, implying that participants became more confident about both types of discriminations (compare Figure 5).

There was also a significant difference in the slopes of the learning rates between e4 vs. e5 ($M_\beta = 0.12$, $SE_\beta = 0.04$) and c3 vs. d3 discriminations ($M_\beta = -0.05$, $SE_\beta = 0.03$), $t(29) = 4.06$, $p < .001$, with the β parameter being significantly different from zero for e4 vs. e5 discriminations, $t(29) = 3.14$, $p = .004$, but not for c3 vs. d3 discriminations, $t(30) = -1.48$, $p = .15$. This indicates that discriminations based on the previously predictive dimension were learned rapidly (i.e., during the first block), whereas those based on the previously nonpredictive dimension had to be learned progressively throughout Stage 2. Together with

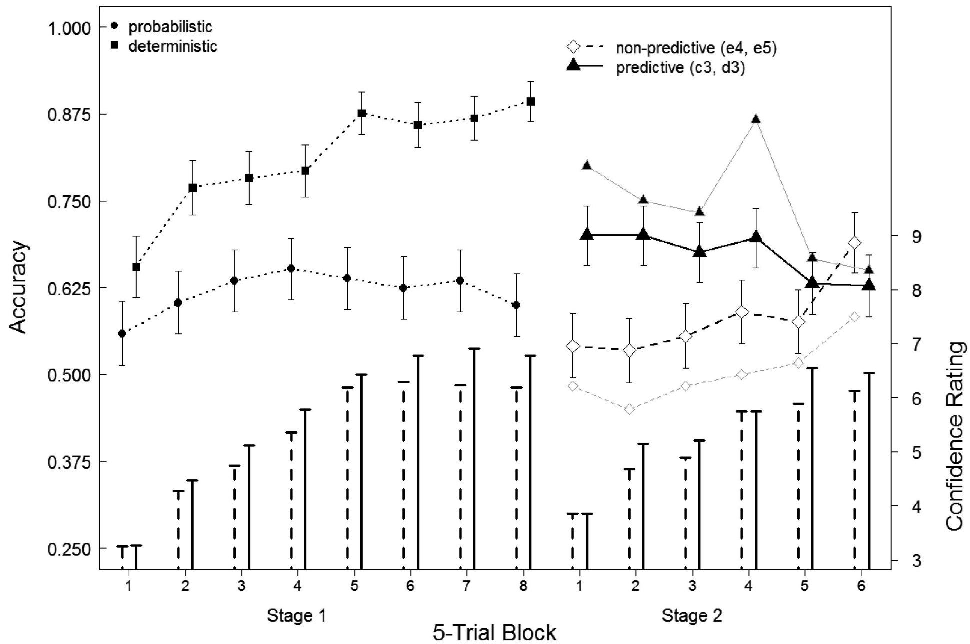


Figure 5. Mean accuracy (left ordinate) and confidence ratings (right ordinate; vertical bars) for the responses across five-trial blocks of Stage 1 (left/right) and Stage 2 (up/down) in Experiment 4. Stage 2 data are shown separately for discriminations based on the previously predictive (solid lines) and nonpredictive (dashed lines) stimulus dimension (the nonpredictive dimension carried information about the uncertainty of the predictions in Stage 1). Error brackets illustrate the standard errors of the mean. The thin lines depict the average accuracy of participants who were presented with less similar values along the previously predictive dimension (see the Results and Discussion, p. 11).

the main effect of predictiveness on accuracy, this pattern of results suggests a positive transfer of dimensional associability (in line with Sutherland & Mackintosh, 1971).

In contrast to a prediction that could be derived from the Pearce and Hall (1980) model, the associability of the uncertainty-indicating stimulus dimension did not increase. Rather, and in line with the Mackintosh (1975) model, the Stage 2 learning rates indicate that the predictive dimension was attended more than the response-irrelevant and uncertainty-predicting dimension. The occurrence of a transfer effect in Experiment 4, as opposed to Experiments 2 and 3, might indicate that a particular cue along the uncertainty-indicating dimension (i.e., Value 2) acted as an inhibitor of the associated response. Generalization of response inhibition to different values along the same (nonpredictive) dimension would thus account for (or have contributed to) the reduced learning rate in Stage 2.

Just as in Experiment 3, a significant Similarity × Predictiveness interaction on accuracy, $F(2, 28) = 3.61$, $p = .040$, $\eta^2_G = 0.04$, indicates that the learned-predictiveness effect in Experiment 4 was moderated by the specific similarity relations of stimulus values (intermediate orientations, intermediate frequencies, or more extreme frequencies in Stage 2 compared with Stage 1). Six participants with spatial frequency being predictive in Stage 1 were presented with more extreme values along this previously predictive dimension in Stage 2 (i.e., 1 vs. 5 cpd). These participants' subsequent learning rates are illustrated as separate thin lines in Figure 5. A 6 (block) × 2 (predictiveness) repeated-measures ANOVA on the accuracy of their responses at Stage 2 revealed a significant main effect of prior predictiveness, $F(1, 5) = 26.56$, $p = .004$, $\eta^2_G = 0.47$, indicating that discriminations were learned faster for the predictive dimension ($M = 0.74$) than for the nonpredictive dimension ($M = 0.50$). There was no main effect of block, $F(5, 25) = 0.57$, $p = .72$, and only a marginally significant interaction, $F(5, 25) = 2.49$, $p = .06$.[2]

In contrast, for participants with intermediate frequencies in Stage 2, the ANOVA revealed a marginally significant main effect for block, $F(5, 40) = 2.40$, $p = .05$, $\eta^2_G = 0.10$, but not for predictiveness, $F(1, 8) = 0.38$, $p = .58$, and there was no interaction, $F(5, 40) = 1.89$, $p = .12$. Similarly, intermediate orientation discriminations at Stage 2 did not reveal a significant main effect of predictiveness, $F(1, 15) = 0.51$, $p = .48$, even though there was an interaction with block, $F(5, 75) = 4.29$, $p = .002$, $\eta^2_G = 0.08$, suggesting differential learning rates for discriminations along the previously predictive and nonpredictive dimension. There was no main effect of block for these participants, $F(5,$

[2] Note that when spatial frequency was irrelevant in Stage 1, the average accuracy for the same 1 versus 5 cpd discrimination in Stage 2 was 0.67 (Subject Nos. 11, 25, 26, 28, and 34; see Appendix C). An additional 2 (relevant dimension: spatial frequency, orientation) × 6 (block) mixed-model ANOVA revealed a significant interaction between relevant dimension and block, $F(1, 9) = 6.15$, $p = .03$, $\eta^2_G = 0.16$, indicating that participants with spatial frequency being relevant in Stage 1 learned the 1 versus 5 cpd discriminations faster ($M_{\text{Block 1}} = 0.80$) than participants with orientation being relevant in Stage 1 faster ($M_{\text{Block 1}} = 0.60$). The main effects in this ANOVA were not significant.

75) = 0.11, p = .99. Taken together, this pattern of results implies that (a) most participants in Experiment 4 demonstrated positive transfer of dimensional predictiveness (i.e., faster learning along the predictive dimension), and (b) the transfer seems to be more pronounced if the particular orientations or frequencies did not lie between the specific discriminations that were learned during Stage 1 (indicating a contribution of stimulus-specific associability).

The results of Experiment 4 suggest that the predictiveness of an entire stimulus dimension can, under certain conditions, lead to positive transfer (i.e., faster learning with an intradimensional shift). The dimensional transfer effect might just be a result of the slightly greater number of participants who were presented with more extreme stimuli in Stage 2 along the previously predictive dimension compared with Experiments 2 and 3. However, the occurrence of a dimensional transfer effect in Experiment 4 might be explained in terms of a "hybrid" model of associative learning (Le Pelley, 2004, 2010). According to this model, the associability of a cue depends on the product of attentional *exploitation* (a mechanism based on the Mackintosh model) and attentional *exploration* (an uncertainty-driven attentional search based on the Pearce-Hall model). Consistent with this model, Beesley, Nguyen, Pearson, and Le Pelley (2015) found that attention during associative learning depends on both predictiveness and uncertainty, with uncertainty exerting an effect on attention at a between-compound level (i.e., enhanced attentional "exploration" of cues that belonged to uncertainty-indicating compounds), and the predictiveness of cues determining the relative amount of attention that is directed to cues within a compound (i.e., enhanced "exploitative" attention on the predictive cue of a compound). In similar veins, an entire stimulus dimension that carries uncertainty information might lead to greater exploratory attention, and thus account for the positive transfer effect that was observed in Experiment 4. Specifically, enhanced attentional exploration might have helped participants to discover the relevant dimension and thus have produced dimensional transfer of predictiveness. At the same time, exploitative attention is directed to particular stimulus values within the predictive dimension (e.g., certain spatial frequencies), which are better predictors than other stimulus values. Accordingly, the transfer effect was enhanced for Stage 2 discriminations that were compatible with the previously learned discriminations (i.e., more extreme values along the predictive dimension) than for stimulus values that lay between those that were attended (exploitatively) during Stage 1. Further research is certainly needed to substantiate this suspicion (see also George & Pearce, 2012).

General Discussion

The first experiment reported in the present study replicated the basic learned-predictiveness effect (Le Pelley & McLaren, 2003) in a stimulus–response contingency learning paradigm with perceptually complex Gabor gratings. That is, a higher learning rate (indicating an increase in associability) was found for individual stimuli that were predictive of a response at a previous learning stage compared with previously irrelevant stimuli.

However, in two subsequent experiments, we showed that this positive transfer of predictiveness may not occur in equal measure for the predictiveness of an entire stimulus dimension. On average, the learning rate for test discrimination problems did not depend on whether the distinguishing dimension was previously predictive or nonpredictive. The fact that the associability of a stimulus dimension did not depend on its predictiveness with both deterministic and probabilistic stimulus–response associations indicates that even a higher level of uncertainty during training does not lead to a (positive or negative) transfer of dimensional associability.

In line with many previous studies supporting the Mackintosh (1975) model, Experiment 1 demonstrated faster learning rates for individual Gabor gratings that were learned to be indicative of an outcome over gratings that were poor predictors of an outcome. This indicates that, in line with Equation 4, the associability of individual stimuli increased as a consequence of their higher predictiveness (e.g., A or B) compared with other stimuli (e.g., X or Y). In contrast, when an entire stimulus dimension, rather than individual stimuli, was indicative of the requested response at Stage 1, learning rates did not differ between discriminations that were based on the relevant or the irrelevant dimension. This result is difficult to explain with the assumption that individuals learn to attend to the relevant dimension and to ignore the irrelevant dimension (Sutherland & Mackintosh, 1971). Instead, attention seems to be directed to particular stimuli that are informative, whereas stimuli lying between those stimuli along the predictive dimension are ignored. Specifically, according Mackintosh, the intradimensional shift effect is assumed to be the result of direct transfer of associability depending on similarity, with features from one stimulus dimension typically being more similar to each other than to features of a different dimension. However, this model also predicts a transfer of dimensional predictiveness (i.e., an intradimensional shift advantage) to be absent if the test stimuli "interfere" with the trained discrimination.

Trobalon et al. (2003) showed that the intradimensional shift advantage disappeared (for spatial discrimination learning in rats) when the discriminations at Stage 2 lay between the discriminations that were learned at Stage 1. This indicates that the rats learned to attend to the specific locations (which were relevant for the spatial discriminations), whereas they learned to ignore the irrelevant locations (e.g., those lying between the crucial locations). In Experiments 2 and 3 of the present study, the stimuli used at Stage 2 were drawn randomly, based on a previously either relevant or irrelevant continuous dimension. However, just as in the Trobalon et al. study, the orientations and most of the frequencies that were relevant at Stage 2 lay between those that were relevant to the Stage 1 discriminations. Although attending to the predictive stimuli in Stage 1, most participants may thus have learned to ignore a range of stimuli that would be predictive in Stage 2. Hence, there was no transfer of dimensional predictiveness. For a few participants, however, the stimuli that were presented at Stage 2 were compatible with the Stage 1 discriminations (i.e., more extreme examples of the stimuli that were learned), and these individuals actually learned later discriminations more readily when they were based on the previously predictive dimension. Although these data are difficult to explain with a learning theory that assumes changes in associability of entire stimulus dimensions (Sutherland & Mackintosh, 1971), they are compatible with the assumption that learning implies only changes in the associability of particular stimuli (depending on their predictiveness; Mackintosh, 1975). More precisely, the present data suggest that individuals learn to attend to specific discriminatory stimulus values (e.g., spatial frequencies of 2 and 4 cpd) while ignoring irrelevant information (e.g., the orientation of the grating and spatial frequencies between 2 and 4). With the acquired attentional bias, subsequent discrimination problems along the spatial frequency dimension will be easiest to learn for equal or more extreme spatial frequencies (e.g., 1 and 5 cpd) than for intermediate spatial frequencies (e.g., 1 vs. 3 or 3 vs. 5 cpd). This clearly suggests that discriminations along a previously predictive dimension are not always learned faster than discriminations along an irrelevant dimension (i.e., an intradimensional shift is not always superior to an extradimensional shift). Rather, the rate of learning seems to depend on whether the specific stimuli were attended or ignored in order to learn previous discriminations.

Finally, in Experiment, 4 half of the cues had a deterministic association with a response, whereas the other half of cues was probabilistically associated with a response, with one dimension being predictive and the other dimension being indicative of the uncertainty of the association at Stage 1. Thus, in contrast to the two previous experiments, the response-irrelevant dimension indicated how reliably a stimulus value in the other relevant dimension predicted the outcome, and here we observed a clear transfer effect for dimensional associability. In particular, the learning rate at Stage 2 was enhanced for discriminations that were based on the previously predictive stimulus dimension (orientation or spatial frequency) compared with the irrelevant but uncertainty-indicating dimension (reflecting a positive transfer of dimensional associability in line with Sutherland & Mackintosh, 1971). It might be speculated whether the correlation of an entire stimulus dimension with uncertainty in this experiment led to a change in associability at the level of an entire stimulus dimension. That is, an uncertainty-based learning mechanism (aiming at the correction of prediction errors, as suggested by Pearce & Hall, 1980) could have directed attention to one stimulus dimension in favor of another dimension. Nevertheless, even in Experiment 4, we observed that the transfer of predictiveness also depended on the specific relations between the stimuli presented at training and test. Just as in the previous experiments, the greatest differences in learning rate as a function of prior predictiveness were found when the transfer stimuli did not lie between the trained stimuli (and were thus not to be ignored during training). This clearly indicates that attention had also been directed to particular stimuli (Mackintosh, 1975), even when uncertainty information may have biased the associability of an entire dimension.

In principle, the fact that transfer of dimensional predictiveness was found in Experiment 4 but not in Experiments 2 and 3 could be a result of the slightly greater number of participants who were presented with more extreme stimuli along the predictive dimension. However, the results of Experiment 4 may also be consistent with certain dual-process accounts of associability. According to George and Pearce (2012), for instance, the associability of specific stimuli is assumed to be affected by a predictiveness-driven mechanism (Mackintosh, 1975), whereas the associability of entire stimulus configurations (or stimulus dimensions) might be based on an uncertainty-driven learning mechanism (Pearce & Hall, 1980). In addition, as outlined previously, the transfer of dimensional predictiveness might be based on greater attentional exploration in case of an uncertainty-indicating stimulus dimension that increased the likelihood of detecting the relevant stimulus dimension (and attention being directed to the entire dimension as opposed to specific stimulus values). Simultaneously, and thus in line with a hybrid model, the greater positive transfer for particular test discriminations may be the result of attentional exploitation of the predictive stimulus values within the relevant dimension (see Beesley et al., 2015; Le Pelley, 2004, 2010). Though the present data do not allow us to distinguish between these accounts (this is beyond the scope of this article), the pattern of results we observed in the current series of experiments seems compatible with these ideas.

An obvious limitation of the present study is the fact that the similarity relations between the discriminatory stimuli in the two learning stages were not manipulated systematically. More research is required to verify the assumption that positive transfer of predictiveness will occur if the test discriminations consist of specific stimuli that the individuals learned to attend to during the previous learning stage, whereas previous learning experience will not transfer to stimuli that lie between the trained examples. Future studies can address this issue by contrasting test discriminations with stimuli that are either more or less extreme along the previously predictive or nonpredictive stimulus dimension (keeping the perceptual distances between training and transfer stimuli constant). In line with the present findings (and consistent with a similarity-based account of the changes in associability; Mackintosh, 1975), positive transfer of predictiveness after an intradimensional shift (i.e., faster learning with the previously predictive dimension being relevant) might be expected only for the more extreme test discriminations, but not for test discriminations with intermediate stimuli.

Taken together, we demonstrated that the associability of a predictive stimulus dimension does not necessarily increase compared with a nonpredictive stimulus dimension. That is, in line with previous animal-learning results (Trobalon et al., 2003), we did not find increased learning rates with an intradimensional shift than with an extradimensional shift in many cases in which the specific stimulus relations implied a certain degree of "interference" between training and the transfer stage (i.e., when the discriminatory stimulus values lay between those that were predictive in the previous learning stage). However, when the transfer stimuli were compatible with the discrimination that was learned in an earlier stage (i.e., more extreme stimulus values along the predictive dimension), then faster learning was observed along the previously predictive dimension. These results suggest that learning usually implies changes in associability of individual stimuli (Mackintosh, 1975), but not of entire stimulus dimensions (as suggested by Sutherland & Mackintosh, 1971). Future research is required to address the question of whether a transfer of dimensional predictiveness can be found when uncertainty information leads to greater attentional exploration.

References

Beesley, T., Nguyen, K. P., Pearson, D., & Le Pelley, M. E. (2015). Uncertainty and predictiveness determine attention to cues during human associative learning. *The Quarterly Journal of Experimental Psychology, 68,* 2175–2199. http://dx.doi.org/10.1080/17470218.2015.1009919

Bonardi, C. H., Graham, S., Hall, G., & Mitchell, C. J. (2005). Acquired distinctiveness and equivalence in human discrimination learning: Evidence for an attentional process. *Psychonomic Bulletin and Review, 12,* 88–92.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 443–446. http://dx.doi.org/10.1163/156856897X00357

Bush, R. R., & Mosteller, E. (1951). A mathematical model for simple learning. *Psychological Review, 58,* 313–323.

Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Dissociation of affective and attentional shifting by selective lesions of prefrontal cortex. *Nature, 380,* 69–72.

Esber, G. R., & Haselgrove, M. (2011). Reconciling the influence of predictiveness and uncertainty on stimulus salience: A model of attention in associative learning. *Proceedings*. Biological Science/The Royal Society, *278,* 2553–2561.

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.

George, D. N., & Pearce, J. M. (2012). A configural theory of attention and associative learning. *Learning and Behavior, 40,* 241–254.

Griffiths, O., Johnson, A. M., & Mitchell, C. J. (2011). Negative transfer in human associative learning. *Psychological Science, 22,* 1198–1204.

Griffiths, O., & Le Pelley, M. E. (2009). Attentional changes in blocking are not a consequence of lateral inhibition. *Learning and Behavior, 37,* 27–41.

Hall, G., & Rodriquez, G. (2010). Attentional learning. In C. J. Mitchell & M. E. Le Pelley (Eds.), *Attention and associative learning: From brain to behaviour* (pp. 41–70). Oxford, UK: Oxford University Press.

Hogarth, L., Dickinson, A., Austin, A. J., Brown, C., & Duka, T. (2008). Attention and expectation in human predictive learning: The role of uncertainty. *The Quarterly Journal of Experimental Psychology, 61,* 1658–1668.

Kattner, F. (2015). Transfer of absolute and relative predictiveness in human contingency learning. *Learning and Behavior, 43,* 32–43.

Kaye, H., & Pearce, J. M. (1984). The strength of the orienting response during Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 10,* 90–109.

Kruschke, J. K. (2001). Towards a unified model of attention in associative learning. *Journal of Mathematical Psychology, 45,* 812–863.

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

Le Pelley, M. E. (2010). The hybrid modeling approach to conditioning. In N. A. Schmajuk (Ed.), *Computational models of conditioning* (pp. 71–107). Cambridge, UK: Cambridge University Press.

Le Pelley, M. E., Beesley, T., & Griffiths, O. (2011). Overt attention and predictiveness in human contingency learning. *Journal of Experimental Psychology: Animal Behavior Processes, 37,* 220–229.

Le Pelley, M. E., & McLaren, I. P. L. (2003). Learned associability and associative change in human causal learning. *The Quarterly Journal of Experimental Psychology, 56B,* 68–79.

Livesey, E. J., & McLaren, I. P. L. (2007). Elemental associability changes in human discrimination learning. *Journal of Experimental Psychology: Animal Behavior Processes, 33,* 148–159.

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

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approach for testing interval null hypotheses. *Psychological Methods, 16,* 406–419.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–51). Hillsdale, NJ: Erlbaum.

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

Pearce, J. M., & Mackintosh, N. J. (2010). Two theories of attention: A review and a possible integration. In C. J. Mitchell & M. E. Le Pelley (Eds.), *Attention and associative learning: From brain to behaviour* (pp. 11–39). Oxford, UK: Oxford University Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. http://dx.doi.org/10.1163/156856897X00366

Roberts, A. C., Robbins, T. W., & Everitt, B. J. (1988). The effects of intradimensional and extradimensional shifts on visual discrimination learning in humans and non-human primates. *Quarterly Journal of Experimental Psychology, 40B,* 321–341.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review, 16,* 225–237.

Suret, M. B., & McLaren, I. P. L. (2003). Representation and discrimination on an artificial dimension. *Quarterly Journal of Experimental Psychology, 56B,* 30–42.

Suret, M. B., & McLaren, I. P. L. (2005). Elemental representation and associability: An integrated model. In A. J. Wills (Ed.), *New directions in human associative learning* (pp. 155–187). Mahwah, NJ: Erlbaum.

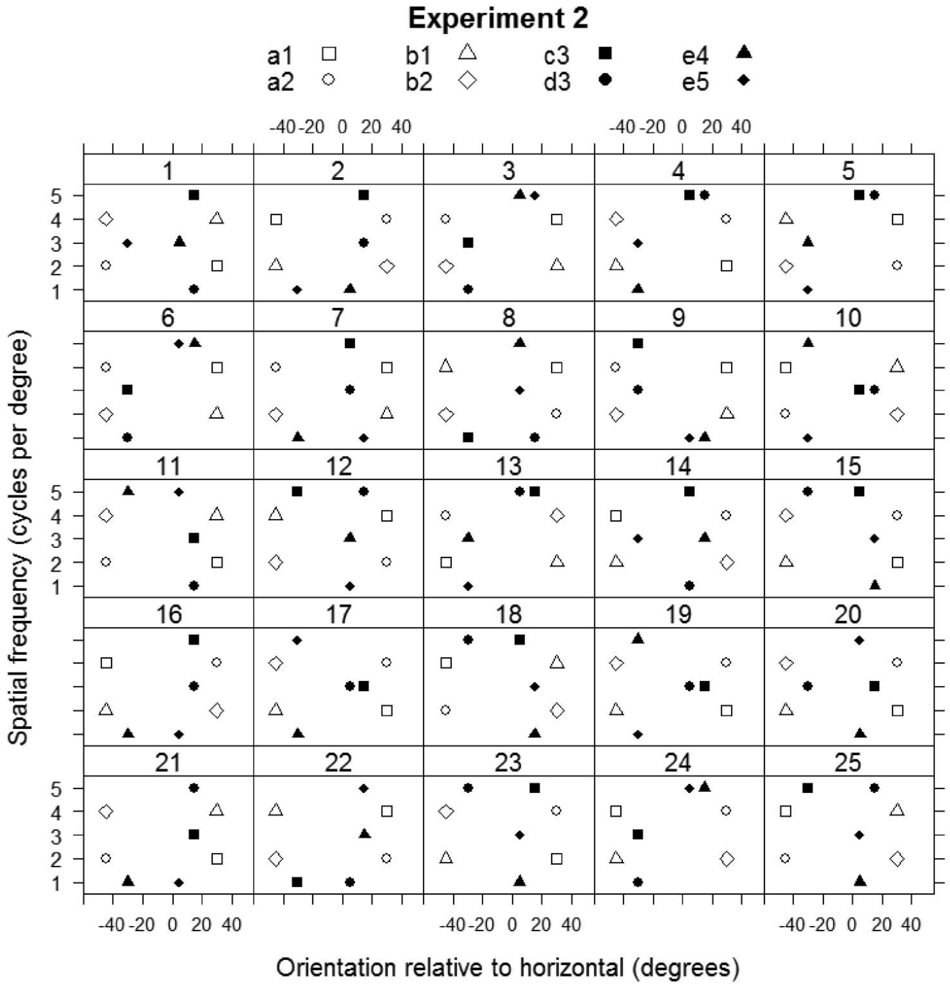
Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. New York, NY: Academic Press.

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.

Whitney, L., & White, K. G. (1993). Dimensional shift and the transfer of attention. *Quarterly Journal of Experimental Psychology, 46B,* 225–252.

Appendix A

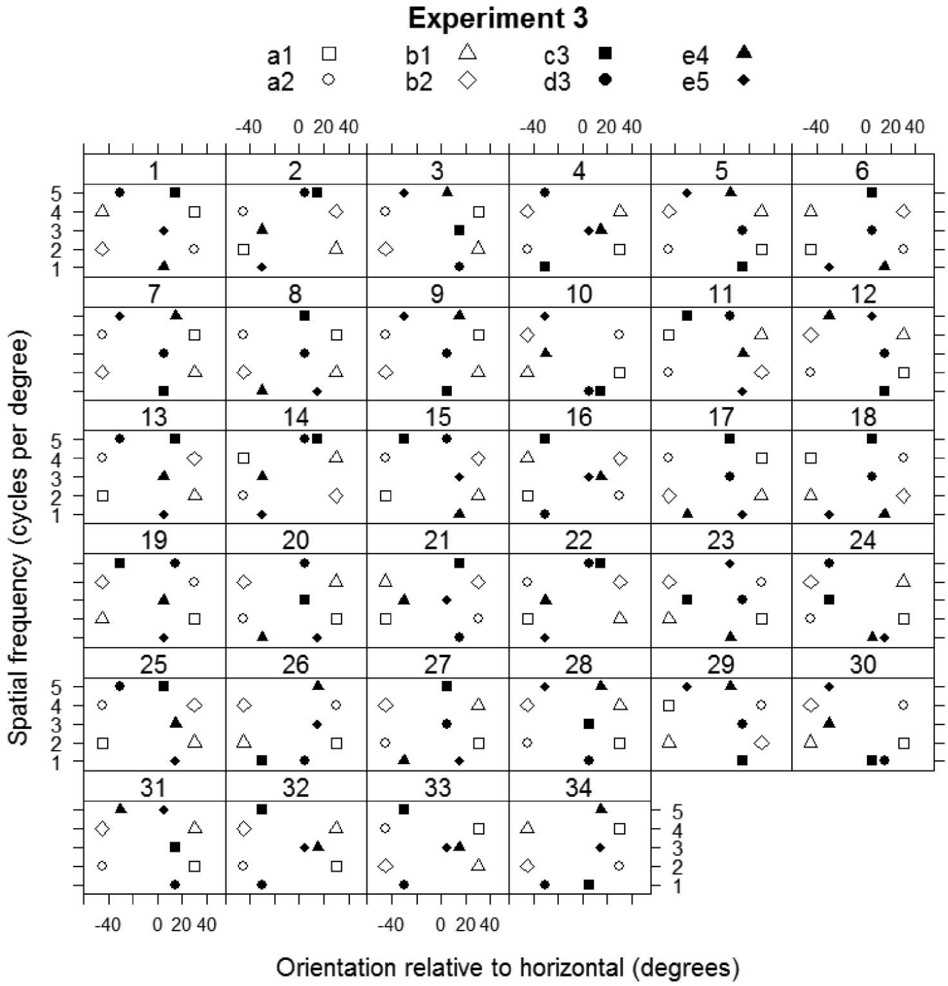
Stimulus Assignments in Experiment 2



(*Appendices continue*)

Appendix B

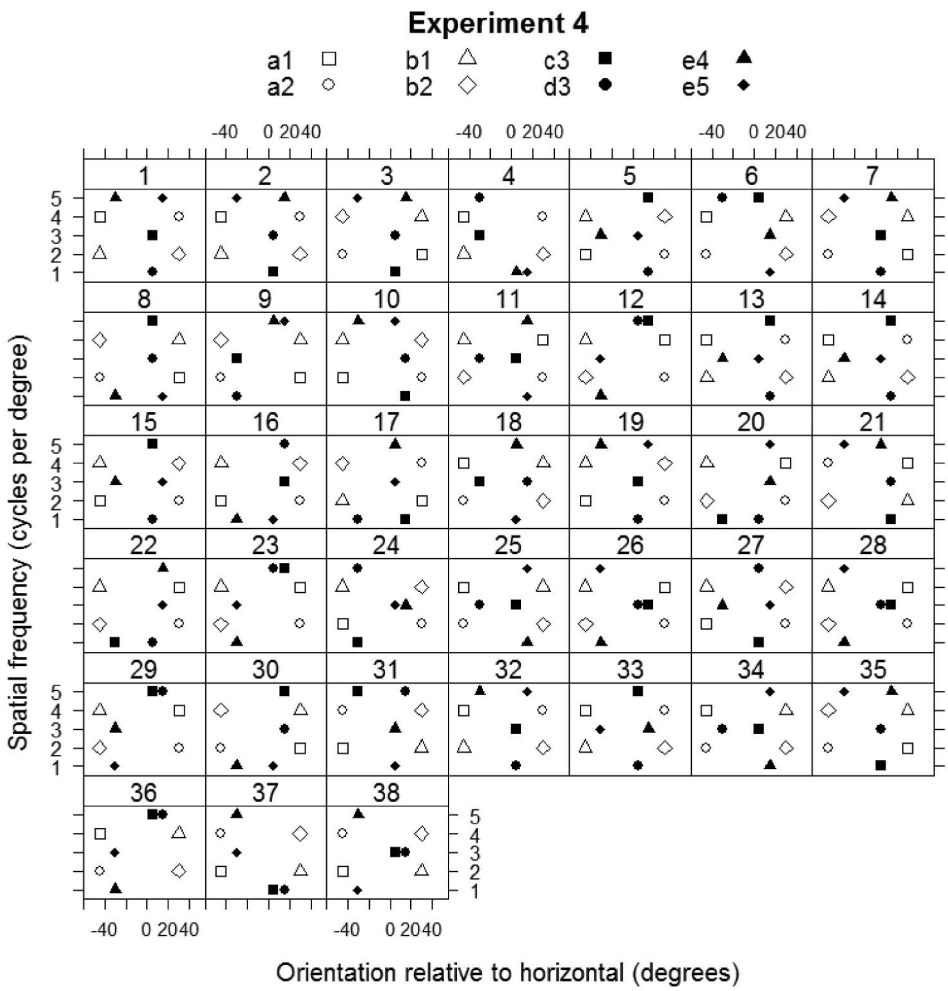
Stimulus Assignments in Experiment 3



(*Appendices continue*)

Appendix C

Stimulus Assignments in Experiment 4



Received March 23, 2015
Revision received August 26, 2015
Accepted August 30, 2015 ■