Mechanisms Underlying the Accuracy of Stimulus Representations: Within-event Learning and Outcome Mediation.

Sandra Lagator[1], Clara Muñiz-Diez[2], Tom Beesley[2], Mark Haselgrove[1]

[1]School of Psychology, The University of Nottingham, Nottingham, UK

[2]Department of Psychology, Lancaster University, Lancaster, UK

Author Note.

Please address correspondence to: Sandra Lagator, School of Psychology, University of Nottingham, University Park, Nottingham, NG7 2RD, UK. Email: Sandra.Lagator@Nottingham.ac.uk. This research was supported by grant ES/W013215/1 from the Economic and Science Research Council of the UK.

Abstract

Valid predictors of an outcome attract more attention than stimuli which are non-predictive. Furthermore, stimuli which have a probabilistic association with an outcome attract more attention than stimuli which have a deterministic association with an outcome. Two experiments investigated whether predictive validity and outcome uncertainty resulted in the establishment of a *more accurate* stimulus representation, in which accuracy was measured as the strength of associations between different elements of a compound stimulus. In Experiment 1, pairs of stimuli were established as outcome-predictive (always followed by the same outcome), and presented in conjunction with non-predictive pairs of stimuli (equally likely to be followed by two different outcomes). Outcome uncertainty was also manipulated, between groups, by establishing either a deterministic (100%) or probabilistic (80%) contingency between the predictive pairs and their outcomes. Test trials revealed more accurate recognition for which predictive stimuli were paired together relative to non-predictive stimuli; however, there was no effect of outcome uncertainty. Experiment 2 reproduced the effect observed in the deterministic group from Experiment 1 and also demonstrated that the superior performance to the predictive stimuli over the non-predictive stimuli was only evident when, at test, the choice stimuli had predicted different outcomes during training. These results were interpreted as the consequence of two pathways to accurate stimulus representation: direct (within compound associations) and indirect (mediated through the activation of the outcome), and discussed in the context of attentional theories of associative learning.

Keywords: Attention; Learning; Within-compound; Representation; Outcome mediation; Uncertainty

Introduction

An enduring question within the study of learning, both in human and non-human animals, is how the processing of a stimulus changes as a consequence of variations in the relationship between it and a subsequent outcome. Two general approaches have emerged to address this question, and both have become associated with the idea that learning can have an impact upon the amount of attention that is paid to a stimulus (Le Pelley et al., 2016). On the one hand there is a proposal that attention to a stimulus will change as a function of its predictive validity for an outcome; on the other hand, there is a proposal that attention to a stimulus will change through variations in outcome uncertainty. The first approach is perhaps best typified by the theory proposed by Mackintosh (1975; see also: Kruschke, 2001, 2003; Le Pelley, 2004), who suggested that stimuli which are relatively good predictors of an outcome are afforded more attention compared to stimuli that are relatively poor predictors of an outcome. According to Mackintosh's model, for example, if a stimulus is the best predictor of the outcome on a trial it will enjoy a gain in attention; if, however, the stimulus is not the best predictor of the outcome then it will suffer a decrease in attention. Therefore, across a series of learning trials, stimuli that have relatively good predictive validity of motivationally important or task-relevant outcomes should capture more attention compared to stimuli that are predictively redundant. This notion has been likened to a process of *exploitative attention*, as it is advantageous to shift focus towards a task-relevant stimulus in order to better exploit this knowledge for gains (Beesley et al., 2015). On the other hand, Pearce and Hall (1980; see also: Le Pelley et al., 2012; Pearce et al., 1982; Schmajuk et al., 1996) proposed an alternative account of the relationship between learning and attention. According to their model, the processing of a stimulus is superior when it has uncertain, rather than predictable consequences. Pearce and Hall suggested that stimuli which are followed by unpredictable outcomes have higher "effectiveness" relative to stimuli that have a consistent association

with an outcome. More specifically, the effectiveness of a stimulus was suggested to be proportional to the size of the prediction error experienced on the previous trial. This notion has been likened to a process of *exploratory attention*, as organisms should dedicate more processing resources to explore stimuli whose consequences are unknown in order to identify environmental contingencies (Beesley et al., 2015).

Predictive Validity.

Studies of the first, exploitative, form of learned attention with human participants have frequently used stimulus associability as a proxy for attention, and these studies have shown that stimuli that have a history of being predictive of an outcome subsequently acquire associations with novel outcomes more quickly than stimuli which have a history of being non-predictive. These studies have also taken more direct measures of overt attention, such as eye-gaze. For example, a study by Le Pelley et al. (2011) tested 1) whether predictive stimuli have higher associability relative to non-predictive stimuli, and 2) whether participants spent longer looking at the predictive relative to non-predictive stimuli. On every trial, participants were presented with compounds in which one stimulus was predictive of one of two outcomes that followed, while the other stimulus was non-predictive. In a subsequent stage, the stimuli were arranged into novel compounds of two stimuli, and participants were asked to learn associations with two novel outcomes. Each compound consisted of a previously predictive and a previously non-predictive stimulus in the second stage but, importantly, all the stimuli were now equally predictive of the new outcomes. If the previously predictive stimuli had higher associability by the end of stage 1, then learning should be biased towards these stimuli, and away from the previously non-predictive stimuli during stage 2. A final test confirmed this prediction – the prior predictive validity of a stimulus influenced new learning (see also: Le Pelley & McLaren, 2003; Lochmann & Wills, 2003). Le Pelley et al. also

demonstrated a consistent pattern with overt attention – during phase 1 and phase 2, dwell times were longer to the predictive relative to the non-predictive stimuli and, interestingly, the magnitude of this overt bias in attention towards the predictive stimuli correlated with the bias seen in learning towards the predictive stimuli. In studies conducted with non-human animals, many experiments have used an intra-dimensional/extra-dimensional shift procedure to show that predictive stimuli have higher associability than non-predictive stimuli (e.g., Durlach & Mackintosh, 1986; George & Pearce, 1999; Oswald et al., 2001; Shepp & Eimas, 1964; Roberts et al., 1988; Mackintosh & Little, 1969).

Further studies of predictive validity with human participants have used different measures of attention and have also provided support for the idea that stimuli with high predictive validity acquire relatively high attention (e.g., Haselgrove et al., 2016.; Le Pelley et al., 2013). For example, Le Pelley et al. (2013) used a dot-probe task as a measure of attention in which participants were asked to alternate between two tasks, a category-learning task and a dot-probe task. During the category learning task, two different colours and two different orientations were arranged in colour-orientation compounds. One of the dimensions (e.g., colour) was predictive of the correct response while the other was non-predictive. During the dot-probe task, participants were presented with the same compounds. On some of the trials a dot appeared in the location previously occupied by one of the stimuli (either a colour or an orientation), and participants were asked to press a key whenever they saw the dot-probe. Participants were faster to respond to the probe when it appeared in the location of the previously predictive stimulus relative to when it appeared in the other location, indicating that participants were more likely to attend to the stimulus that was predictive during the category learning task (see also Livesey et al., 2009, for a conceptually similar effect using the attentional blink phenomenon).

Outcome Uncertainty.

An early motivation for the idea that attention declines to a stimulus that is followed by a predictable outcome came from a study by Hall and Pearce (1979) who showed that rats learned more slowly about the relationship between a tone and a strong shock when the tone had previously been established as a reliable predictor for a weak shock, compared to when it was novel. The logic here was that if attention to a stimulus is diminished then learning will be subsequently attenuated (see also: Ayres et al., 1984; Rodriguez & Alonso, 2011; Savastano et al., 1997; Swartzentruber & Bouton, 1986). Interestingly, this effect was attenuated (conditioning could be restored) if two nonreinforced presentations of the stimulus were interposed between the training stages with the weak shock and the stronger shock. The conclusion drawn here was that these non-reinforced trials generated outcome uncertainty and restored attention to the stimulus (Hall & Pearce, 1982; see also: Haselgrove et al., 2010).

Studies with humans have provided mixed evidence for the account of learning and attention proposed by Pearce and Hall (1980). On the one hand, there is consistent evidence that participants spend longer looking at stimuli when they are associated with different outcomes. For example, Beesley et al. (2015), presented participants with two sets of trials, certain and uncertain. On the certain trials (AW, AX, BW, and BX) stimuli A and B were perfectly predictive of outcome 1 and outcome 2 (O1 and O2) respectively, while W and X were non-predictive. On the uncertain trials (CY, CZ, DY, and DZ) stimuli C and D were now probabilistic predictors of outcomes O1 and O2; that is, they were followed by the corresponding outcomes on only 70% of the trials (on other trials the alternative outcome was presented). Stimuli Y and Z were, like W and X on the certain trials, nonpredictive of O1 and O2. The pattern of gaze data supported the idea that inconsistent predictors of the outcomes attracted more attention (e.g., Pearce & Hall, 1980): participants spent longer looking at the

compounds that had a probabilistic association with the outcome relative to the compounds that involved deterministic predictors (see also: Easdale et al., 2019; Hogarth et al., 2008; Walker et al., 2019, 2022). In addition, stimuli which are associated with different outcomes (e.g., Cho & Cho, 2021; Ju & Cho, 2023), such as different monetary rewards (Pearson et al., 2024), will better disrupt an ongoing search task than stimuli which are consistent predictors of an outcome. For animal studies that have shown heightened orienting responses under conditions of outcome uncertainty, see Collins et al. (1983); Collins and Pearce (1985); Honey et al. (1987); Kaye & Pearce (1984); Pearce et al. (1985); Swan & Pearce (1988); Wilson et al. (1992).

Studies with human participants that have examined how successful learning is to stimuli that have a history of certainty or uncertainty have, on the other hand, revealed mixed results. The study by Beesley et al. (2015), described above, while successfully demonstrating that dwell times were longer to stimuli that were followed by two different outcomes relative to stimuli consistently followed by a single outcome, also showed that the subsequent associability of the uncertain stimuli was, if anything, less than that of certain stimuli (see also: Kattner, 2015; Le Pelley et al., 2010; Livesey et al., 2011). In contrast, however, a study by Chao et al. (2021) revealed that the associability of stimuli that have total uncertainty (i.e., a stimulus followed on 50% of the trials with O1, and on 50% of the trials with O2) was subsequently learned about more rapidly than stimuli that were always followed by O1 or O2. This pattern of results was obtained when the total number of different uncertain trials was few (four). When the number of different uncertain trials was increased (to eight) however, then Chao et al. observed a similar pattern to that described by Beesley et al.

Representational Accuracy and Within-Compound Associations

Together, studies of learning in human and non-human animals have shown that stimuli which have high predictive validity, as well as those with a probabilistic association with an outcome, acquire more attention than stimuli which are redundant, or which have a deterministic association with an outcome. Not surprisingly, then, so-called hybrid models of learning and attention have been developed to account for this rather paradoxical state of affairs (e.g., Esber & Haselgrove, 2011; Le Pelley, 2004; Pearce & Mackintosh, 2010). Whatever the relative merits of these different theories, they converge onto common ground – they all acknowledge that the processing of a stimulus can improve as a consequence of learning. Uncertainty, for example, has been argued to engage an exploratory state of information gathering, in order to process more features of a stimulus and reduce prediction errors (e.g., Luque et al., 2017). Similarly, selective attention that is brought about from differential predictive validity has been suggested to reflect better processing of predictive stimulus features over non-predictive features (e.g., Griffiths & Mitchell, 2008). These considerations lead us to the motivation and question of central interest within the current study: does the processing advantage that is acquired under conditions of high uncertainty and predictive validity result in a stimulus whose representation is more accurate.

The notion of representational accuracy - and its relation to learning - has been considered in a number of psychological theories (e.g., Hebb, 1949; Hall, 1991), which suggest that accurate representations are formed through the establishment of within-stimulus links: excitatory associations are formed between the various co-active features of a stimulus to establish a representation of its whole. For McLaren and Mackintosh (2000), like Hebb, one way of establishing a more accurate, veridical, representation was through the formation of associations between the elements of a stimulus that happen to be simultaneously sampled

upon its presentation (also see Estes, 1950; Atkinson & Estes, 1962). From trial to trial, the stimulus representation will vary despite the stimulus itself remaining objectively constant. To overcome this variability, it has been suggested that the elements of a stimulus that are sampled on any given trial will become associated with each other (e.g., McLaren & Mackintosh, 2000). If there is some overlap between the elements sampled on each trial, a spread of activation will be acquired from the sampled elements to the larger population of elements that can be activated by the stimulus. It follows from this analysis, then, that the accuracy of a stimulus representation should depend on the associability of the elements that make it up. A tempting question to then ask is: do predictive validity and outcome uncertainty influence the representational accuracy of a stimulus? Extant studies seem to provide the basis for thinking that this will be the case: stimuli which have high predictive-validity attract more attention and possess higher associability than non-predictive stimuli and so we might naturally think that this will extend to their constituent elements – facilitating within-stimulus associations. Similarly, stimuli which have a probabilistic relationship with an outcome attract longer dwell times, and, although it remains to be fully resolved how this translates to heightened associability, the notion of exploratory attention implies a process of vigilance among the elements that comprise a stimulus (e.g., Luque, et al., 2017). Again, then, we might expect this to be a circumstance that promotes within-stimulus association.

In humans, evidence for the formation of associations between simultaneously presented stimuli, has come from studies that have investigated *within-compound associations,* particularly in the in the context of retrospective revaluation, and the method by which these associations have been determined has employed either indirect or direct methods (Welham & Wills, 2011). Typically, the indirect method will first give participants training in which a compound of two stimuli is followed by one of two outcomes, and participants must make predictive judgements about the identity of the outcome during the

compound. Within-compound associations are then subsequently assessed with a recognition test in which previously seen stimulus-compounds and novel stimulus-compounds are presented to participants, and an "old" or "new" response must be made, respectively (e.g., Larkin et al., 1998; Wasserman & Berglan, 1998). With the direct method, participants will again be required to first predict which one of two outcomes will follow a variety of stimulus compounds. However, at test, within-compound associations are assessed by presenting one of the stimuli from a compound and participants are asked to select the stimulus that accompanied it from a choice of correct and incorrect options (Melchers et al., 2004), or by means of a reaction time to that stimulus (Beesley & Shanks, 2012). Using both techniques, these studies have revealed that within-compound associations are established when a compound of stimuli has been followed by an outcome. However, the focus of attention has been on how within-compound associations may modulate stimulus-outcome associations, not the reverse question. It therefore remains unknown how the predictive validity of a stimulus, or its outcome uncertainty, influences the integrity of the within-compound associations and hence the accuracy of the stimulus representation (for studies with non-human animals that have investigated the properties of within-compound associations between simultaneously presented stimuli, see: Rescorla 1980; 1981; Rescorla & Cunningham, 1978; Rescorla & Durlach, 1981; and Rescorla & Freberg, 1978).

Here we report two experiments with human participants that sought to investigate the contribution of predictive validity and outcome uncertainty to within-compound associations. In Experiment 1 we manipulated predictive validity in a manner that established pairs of stimuli as either predictive or non-predictive of an outcome. At the same time, the predictive pairs could have either a deterministic or a probabilistic association with an outcome. In this way we were able to investigate the degree of within-compound association as a function of the predictive validity of the stimuli, and as a function of outcome uncertainty. To anticipate

our results, participants made more accurate responses in their recognition of which stimuli were paired together for predictive than for non-predictive pairs, but outcome uncertainty did not improve within-compound learning.

Experiment 2 investigated two different routes that could have allowed participants better recognition for the predictive relative to the non-predictive pairs. One possibility is that participants could have formed stronger within-compound associations between the elements of predictive relative to non-predictive pairs, a mechanism that would be in line with the previous literature that showed elevated attention towards predictive stimuli. Another possibility is that participants could have relied on their knowledge of the common outcomes associated with the predictive stimuli. Since the elements of each predictive pair were (by definition) associated with the same outcome, participants could have relied on their memory for the common outcomes to indicate which predictive stimuli were paired together. Again, in Experiment 2, pairs of stimuli were established as either predictive or non-predictive of an outcome; however, the question of interest now was whether the apparent within-compound associations were a consequence of direct associations between the stimuli within the compound, or an indirect association between the stimuli and the outcome.

Experiment 1

Experiment 1 comprised two stages: (1) a training stage in which participants were required to learn the predictive relationships between a compound of stimuli and an outcome and (2) a test stage in which one stimulus from a training compound was presented and participants were required to identify its associate from correct and incorrect options. During the training stage, participants were presented with four different stimulus compounds: ABWX, ABYZ, CDWX, and CDYZ. For Group Certain, stimulus pairs AB and CD were perfectly predictive of O1 and O2 respectively, while stimulus pairs WX and YZ were non-

predictive of the outcomes (i.e., ABWX – O1, ABYZ – O1, CDWX – O2, CDYZ – O2). Our prediction was that, as training progresses, the within-compound associations that develop between the predictive pairs (A and B; C and D) would be stronger than the within-compound associations that develop between the non-predictive pairs (W and X; Y and Z). Group Uncertain was shown the same types of stimulus compounds during the training stage. However, this group experienced outcome uncertainty (Beesley et al., 2015) and thus trials that involved AB were followed by outcome O1 on 80% of the trials, while outcome O2 occurred on 20% of the trials. The reverse was true for the compounds that involved the CD pair (80% O2, 20% O1). Our expectation was that within-compound associations would be stronger overall in Group Uncertain relative to Group Certain (see Table 1), based on the evidence showing that attention is elevated towards stimuli that have a probabilistic association with an outcome. In order to assess the accuracy of the within-compound associations following this training, participants were given recognition tests, similar to the method employed by Melchers et al. (2004). Here, on each trial, participants were presented with a target stimulus (e.g., A) and asked to identify which stimulus from one of two options accompanied it during the previous training stage (e.g., B or C). The predictive and the non-predictive stimuli were tested on separate test trials, so that the predictive trials involved every possible combination of the predictive cues (e.g., A with B vs C, A with B vs D), and the same was true for the non-predictive cues. If predictive validity and outcome uncertainty modulate the accuracy of stimulus representations, then recognition accuracy at test should be better for the predictive stimuli relative to the non-predictive stimuli and, overall, better in Group Uncertain than Group Certain.

Method

Transparency and Openness.

In this study we detail the processes for identifying sample sizes and data exclusions; we also report measures of recognition-test reliability. Frequentist statistical analyses were conducted using R version 4.2.3 (R core team, 2023), and Bayesian analyses were conducted in JASP (version 0.17.2.1) using the default prior. All experiments were built with the open-source software, PsychoPy (Pierce et al., 2019; version 2023.1.3), and all experiments were run on Pavlovia (https://pavlovia.org/). Participants were recruited through Prolific (www.prolific.com). The design and analysis of the experiments were based on previously published manuscripts but were not preregistered. Data, analysis scripts, and materials are freely available at: https://doi.org/10.17605/OSF.IO/6QVKM. All the experiments reported in this paper received ethical approval by the Ethics Committee at the School of Psychology, University of Nottingham, UK.

Participants.

Studies of differences in the associability of predictive vs non-predictive stimuli reveal effect sizes ranging from small-to-medium (η_p^2 = .05, e.g., Thorwart et al., 2017) to medium-to-large (η_p^2 = .10, e.g., Haselgrove et al., 2016). Studies which have revealed that the associability of uncertain cues is superior to that of deterministic cues have revealed medium-to-large, and large effect sizes (η_p^2 = .12 and .27, Experiments 2a and 2b, Chao et al 2021.). We thus erred on the side of caution and aimed to recruit around 36 participants per group. G*Power 3.1 (Faul et al., 2009) revealed that this would provide a power of .90 to detect an interaction with an effect size of η_p^2 = .05, which would reflect a difference in within-compound knowledge about predictive and non-predictive stimuli that varies as a function of experimental group. Eighty-four participants from the UK, USA and Australia

who reported having English as a first language were recruited via Prolific. Participants were excluded if they did not complete the experiment, or if they made more than one attempt at beginning the training stage. Following exclusions, there were 70 participants in total - 35 in Group Certain and 35 in Group Uncertain. Across the experiment, 23 participants identified as females, 30 participants identified as males, and 17 participants chose not to respond. Sixteen participants did not report their age, the age range of the remaining participants was 18-30 years (M = 26.1, SD = 3). The experiment took approximately 40 minutes to complete. Participants were given a £9 inconvenience allowance for their participation.

Table 1

Design of Experiment 1

Training phase				**Test phase**	
Group	Compound	Outcome Probability		Trials	
		O1	O2	Predictive	Non-predictive
Certain	ABWX	1	0	A: **B** vs C	W: **X** vs Y
	ABYZ	1	0	A: D vs **B**	W: Z vs **X**
	CDWX	0	1	B: **A** vs C	X: **W** vs Y
	CDYZ	0	1	B: D vs **A**	X: Z vs **W**
Uncertain	ABWX	0.8	0.2	C: **D** vs A	Y: **Z** vs W
	ABYZ	0.8	0.2	C: B vs **D**	Y: X vs **Z**
	CDWX	0.2	0.8	D: **C** vs A	Z: **Y** vs W
	CDYZ	0.2	0.8	D: B vs **C**	Z: X vs **Y**

Note. Bold typeface indicates the correct option on each test trial.

Stimuli and Apparatus.

Eight different images of cluttered rooms were employed as stimuli A to D and W to Z (see: Appendix 2 for all images used). Each image was produced using DALL·E 1, a large language system that generates realistic images based on a text description (e.g., "a very cluttered bedroom"). For each participant, each of the eight images was assigned, at random, to one of the stimuli A-D and W-Z. PsychoPy (Peirce et al., 2019) was used to present stimuli and control the experimental events. Participants were instructed that they can use their desktop devices to complete the experiment, but not phones or tablets.

Design and Procedure.

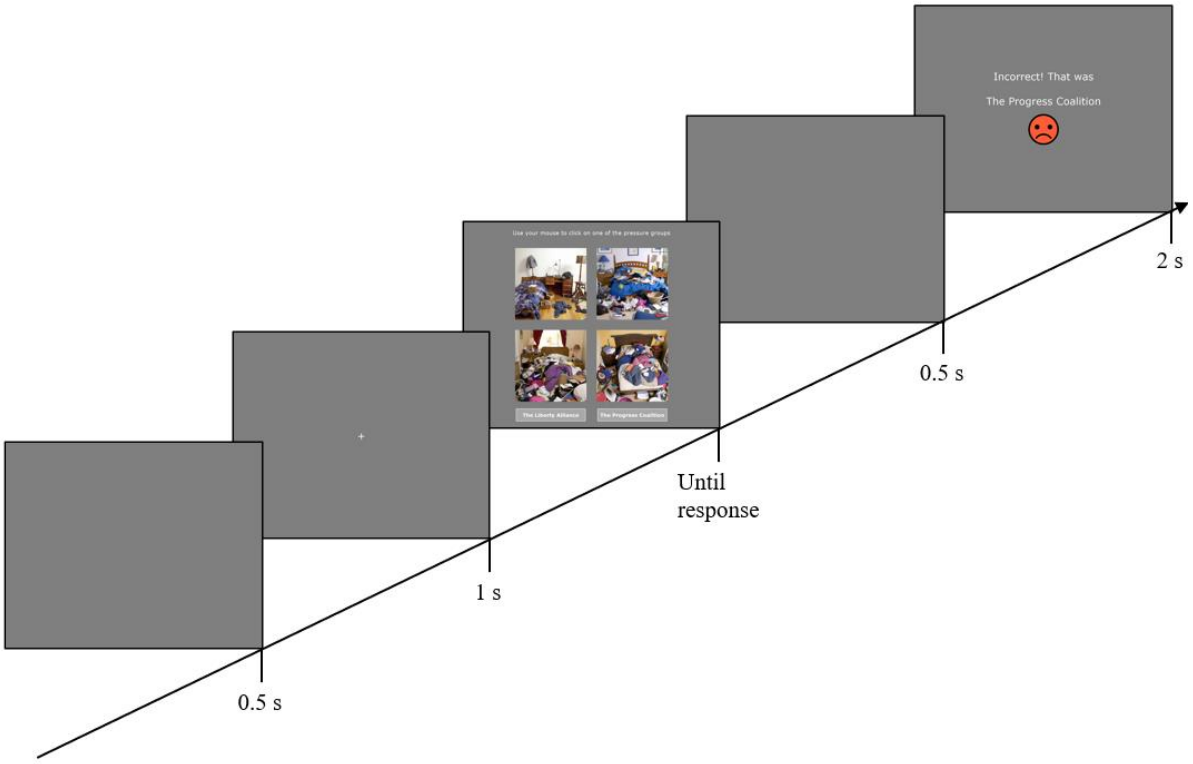
Training Phase. Participants were instructed that they would be engaging in a 'spy training activity'. They were told that two pressure groups (representing O1 and O2), the "Liberty Alliance" and the "Progress Coalition", had begun espionage activities. They were told that these pressure groups had started collecting images of rooms (representing the different stimuli) and that emails had been intercepted that contain these images. Participants were told that each email contained four images of rooms, and that their spy training task was to learn to link these images to the correct pressure group. See Appendix 1 for verbatim instructions.

Training trials began with a grey screen presented for 0.5 s, followed by a fixation cross presented at the centre of the screen for 1s. Participants were then presented with a display showing four different images of rooms (representing four different stimuli, e.g., ABWX). Each image had a size of 0.35 x 0.35 (w x h) in PsychoPy height units, where the size of images is scaled relative to the height of the monitor window. The centres of each image had the following positions, left to right, top to bottom: (-0.2, 0.2), (0.2, 0.2), (-0.2, -

0.2), (0.2, -0.2), where the bottom left corner of the screen window had the location (-0.8, -0.5), the top right corner had the position (+0.8, +0.5) and the centre is (0, 0). The position of each image was randomly determined to one of these four locations on each trial. Below the images, were two buttons, one for each of the pressure groups representing outcomes O1 and O2 (the Liberty Alliance & the Progress Coalition). Each outcome button had a size of 0.34 x 0.06. The centres of the left and right outcome buttons were positioned at (-0.2, -0.44) and (0.2, -0.44) respectively. There was also a statement at the top of the screen instructing participants to use their mouse to click on one of the pressure groups; this display remained on the screen until participants made their response. Following a response, the display was cleared and replaced with a grey screen for 0.5 s, followed by a feedback statement informing participants whether their response was correct/incorrect, and which response was the correct outcome ('Correct/Incorrect! That was the Liberty Alliance/the Progress Coalition'). This display also showed a green smiley face on correct trials, and a red sad face on incorrect trials. The next trial then commenced. See Figure 1 for an example training trial.

Figure 1

An example of a Training Trial in Experiment 1



Note. The timings represent the duration of each display.

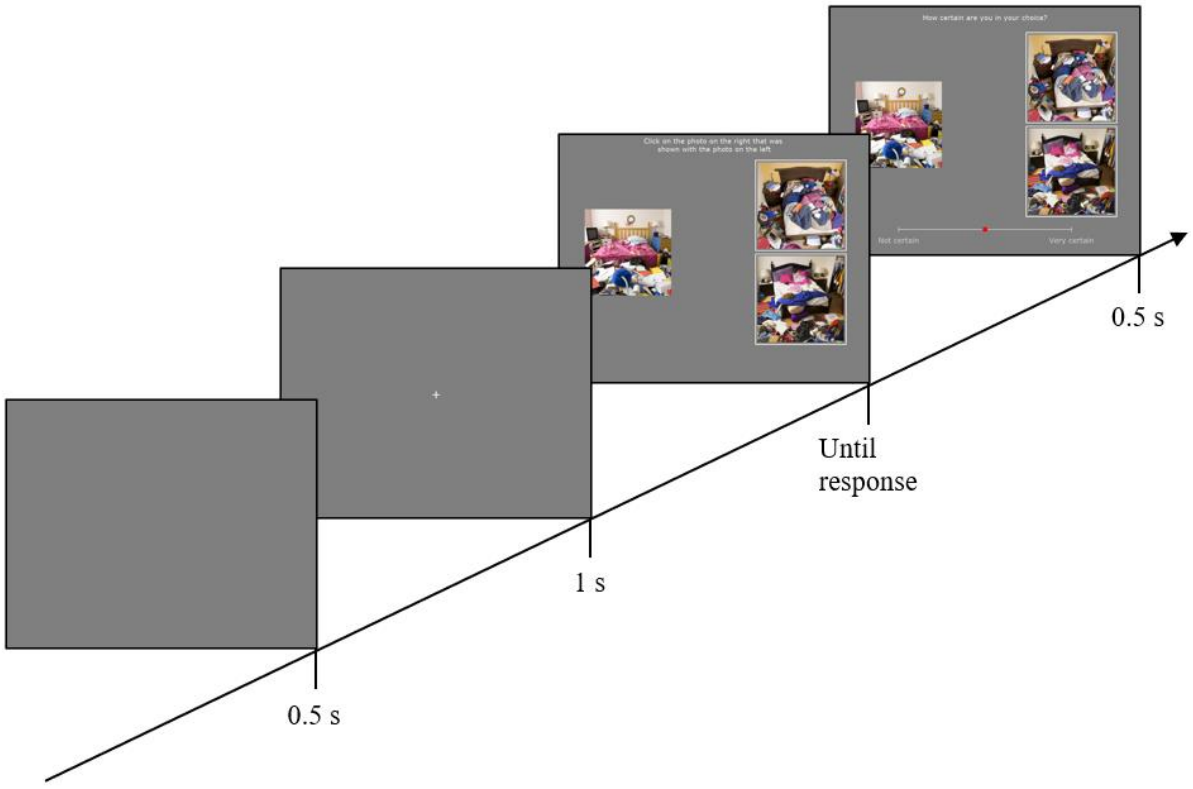
Each block of the training phase consisted of eight trials in total, two for each trial type (ABWX, ABYZ, CDWX, CDYZ); the trials within a block were randomly ordered. Participants completed 20 training blocks, generating 160 trials in total. In Group Certain, the relationship between the predictive stimuli and the outcomes was deterministic – whenever a trial included stimulus pair AB, the outcome that followed was O1; whenever a trial involved stimulus pair CD, the outcome that followed was O2. For Group Uncertain, the predictive stimuli had a probabilistic relationship with the outcome: AB was followed by O1 on 80% of the trials and CD was followed by O2 on 80% of the trials; on the rest of the trials, AB was followed by O2, and CD was followed by O1. Therefore, during the training phase in the

Uncertain condition, there were eight infrequent trials for each trial type (ABWX – O2, ABYZ – O2, CDWX – O1, CDYZ – O1). The positions of the infrequent trials during the training phase in Group Uncertain were randomly determined at the start of the experiment.

Test Phase. Following the training phase, participants were told that the next task will test their memory for the photos seen during the first phase (see Appendix 1 for verbatim instructions). On every trial of the recognition test, participants were presented with a target image on the left side of the screen, and two other images on the top right and bottom-right of the screen, one of which was paired with the target during the training phase and one which was not (see Figure 2 for an example test trial). The predictive and the non-predictive stimuli were presented on separate test trials since all the predictive stimuli were paired with all the non-predictive stimuli during the training phase (e.g., AB was paired with both WX and YZ, thus a non-predictive stimulus could never be an incorrect option for a predictive target stimulus). Each stimulus (A, B, C, D, W, X, Y, Z) represented the target image on two different trials during the test, once for each possible incorrect option (e.g., if the target was A, the correct option was B, while the incorrect option could be C on one trial and D on another trial). Each combination of test stimuli was presented twice to counterbalance the positions of the correct and the incorrect option, e.g., 'A with B vs C' and 'A with C vs B'. Therefore, every stimulus represented the target stimulus on four different trials, which generated 32 trials in total. All the test trials were presented in a random order.

Figure 2

An Example of a Recognition Task Trial in Experiment 1



Note. The timings represent the duration of each display.

Each test trial started with a grey screen shown for 0.5 s, followed by a white fixation cross presented in the centre of the grey screen for 1 s. Participants were then presented with a display showing three different images: a target image and two other images, one of which was paired with the target during the training phase. The images were the same size as used during the training phase (0.35 x 0.35). The target image was shown on the left side of the screen (left image centre at (-0.35, 0.025)); while the two options were shown on the right side of the screen, the top image centred at (0.35, 0.215), and the bottom image centred at (0.35, -0.165). There was a white rectangle surrounding each of the response options. At the top of the screen, there was a statement informing participants to click on one of the two

photos on the right that was shown with the photo on the left during training. This display remained on the screen until participants selected one of the response options. A response was immediately followed by a display asking participants to provide a confidence rating. The three images remained on the screen; the statement at the top changed to 'How certain are you in your choice?', and a slider appeared at the bottom of the screen. The slider had seven increments with labels 'Not certain' and 'Very certain' shown at the left- and right-hand ends of the slider respectively. There was a red circle in the middle of the slider indicating the starting position. Participants could click on the scale or drag the red circle to a chosen position; the trial ended when they released the mouse button.

Results

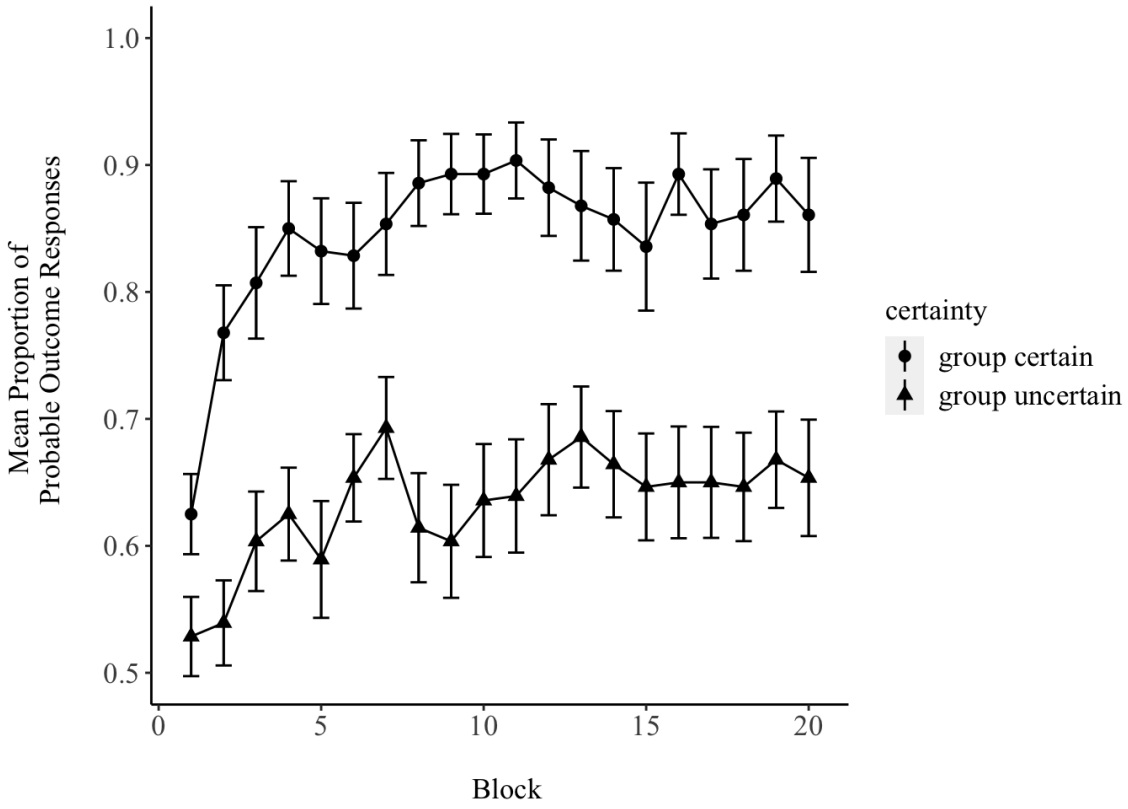
In all statistical tests, we adopt a significance level of .05. Greenhouse–Geisser corrected degrees of freedom were used where Mauchly's test indicated that the assumption of sphericity was violated.

Training accuracy. Participants in Group Uncertain could not make a correct response on every trial even if they always guessed the more likely outcome (e.g., responding O1 on ABWX or ABYZ trials). Therefore, we used the proportion of probable outcome responses as a measure of accuracy (i.e., if participants guessed the more likely outcome, irrespective of the identity of the outcome on that individual trial, this counted as a correct answer; Beesley et al., 2015). For each participant, the proportion of probable outcome responses was calculated individually for each training block, and Figure 3 shows the mean accuracy across the 20 training blocks for Groups Certain and Uncertain. There was an increase in response accuracy across the training blocks for both groups, but overall accuracy was lower for Group Uncertain than Group Certain. These impressions were confirmed with a mixed 2-way Analysis of Variance (ANOVA) of individual proportions of probable

outcome responses with the between-subjects factor of group (certain & uncertain) and block (1-20) as a within-subjects factor. This revealed significant main effects of block, $F(10.75, 730.79) = 6.82$, $p < .001$, $\eta_p^2 = .09$, and of group, $F(1, 68) = 24.31$, $p < .001$, $\eta_p^2 = .26$. The interaction between group and block was not significant, however, $F(10.75, 730.79) = 1.36$, $p = .191$, $\eta_p^2 = .02$. The proportion of probable outcome responses during the last block of training was above chance (0.5) in both Group Certain, $t(34) = 8.03$, $p < .001$, $d = 1.36$, and Group Uncertain, $t(34) = 3.35$, $p = .002$, $d = 0.57$.

Figure 3

Training Accuracy in Experiment 1: The Proportion of Probable Outcome Responses across the Training Blocks in Groups Certain and Uncertain



Note. Error bars represent +/-1 standard error of the mean.

Recognition test accuracy. The crucial data for Experiment 1 were collected during the test phase, where we sought to determine (1) whether the accuracy of within-compound associations between the predictive stimuli was higher than for the non-predictive stimuli and (2) if, overall, this accuracy was higher for Group Uncertain than Group Certain. For each participant, the proportion of correct stimulus selections at test was calculated separately for the predictive and the non-predictive test trials, and Figure 4 shows the means of these proportions. For Group Certain, but not Group Uncertain, recognition accuracy was higher for the predictive stimuli relative to the non-predictive stimuli. Furthermore, overall, accuracy was comparable between Groups Certain and the Uncertain. A two-way mixed ANOVA with the between subjects factors of group (Certain vs. Uncertain) and predictiveness (predictive & non-predictive test trials) showed that the main effect of group was not significant, $F(1, 68) = 0.25$, $p = .622$, $\eta_p^2 = .004$, and there was no main effect of predictiveness, $F(1, 68) = 3.38$, $p = .070$, $\eta_p^2 = .05$. However, there was a significant interaction between group and predictiveness, $F(1, 68) = 6.01$, $p = .017$, $\eta_p^2 = .08$.

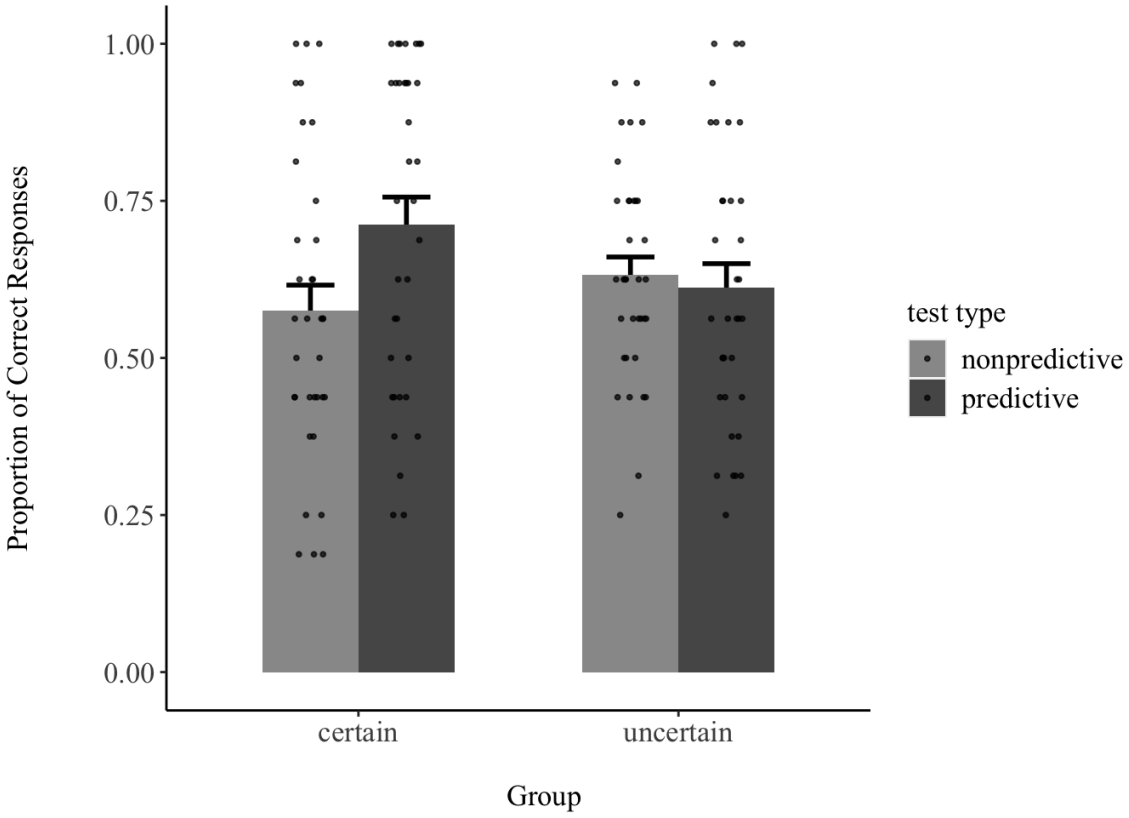
Two one-way ANOVAs showed that participants had higher accuracy on the predictive relative to the non-predictive test trials in Group Certain, $F(1, 34) = 8.89$, $p = .005$, $\eta_p^2 = .21$, but not in Group Uncertain, $F(1, 34) < 1$, $p = .662$, $\eta_p^2 = .006$. Given the importance of the effect of group to our theoretical analysis we conducted a Bayesian repeated-measures ANOVA with the same factors as in the previous frequentist ANOVA; our interpretations for different magnitudes of Bayes factor are based on Andraszewicz et al. (2014). This analysis produced consistent patterns: there was moderate support for the null for the effect of group, $BF_{Inclusion} = 0.26$; a lack of evidence to support either the null or the alternative hypothesis for the effect of predictiveness, $BF_{Inclusion} = 0.78$, while there was moderate support for the interaction between uncertainty and predictiveness, $BF_{Inclusion} = 3.54$. In Group Certain, there was moderate support for the difference in accuracy between the predictive and the non-

predictive test trials, $BF_{Inclusion}$ = 8.9, but moderate support for the null in Group Uncertain, $BF_{Inclusion}$ = 0.28. We also investigated the effect of group at each level of predictiveness. Two one-way ANOVAs showed that the effect of group was not significant for the predictive test trials, $F(1, 68) = 3.03$, $p = .086$, $\eta_p^2 = 0.04$, nor for the non-predictive test trials, $F(1,68) = 1.31$, $p = .256$, $\eta_p^2 = 0.02$. Two Bayesian one-way ANOVAs showed anecdotal support for the null for the predictive test trials, $BF_{Inclusion}$ = 0.89, as well as for the non-predictive test trials, $BF_{Inclusion}$ = 0.43.

For Group Certain, the proportion of correct stimulus selections was above chance (0.5) for the predictive, $t(34) = 4.89$, $p < .001$, $d = .83$, but not for the non-predictive test trials $t(34) = 1.83$, $p = .075$, $d = 0.31$. For Group Uncertain, the proportion of correct stimulus selections was above chance for both the predictive and non-predictive test trials, $ts(34) > 3$, $ps < .005$, $ds > 0.51$, (Bonferroni corrected critical $ps = .0125$).

Figure 4

Recognition Test Accuracy in Experiment 1: The Proportion of Correct Stimulus Selections during the Predictive and the Non-predictive test Trials in Groups Certain and Uncertain



Note. Bars represent mean proportion of correct responses; points represent proportion correct responses for individual participants; error bars represent +/-1 SE.

A Shapiro-Wilk test indicated that the proportions of correct responses had a significant deviation from a normal distribution, $W = 0.95$, $p < .001$. We therefore re-analysed the recognition accuracy data using non-parametric alternatives to investigate the reliability of the preceding analysis. To explore the interaction (the difference between the differences) a difference score for the proportion of correct responses between the predictive and the non-predictive test trials was computed for each participant. A Mann-Whitney U test revealed that there was a significant difference in the difference scores between the two groups, $W = 816$, $p = .017$. Given that there were ties in the ranking of the difference scores, we also ran permutation tests to determine how often the observed test statistics might occur by chance. The individual difference scores were randomly allocated to Group Certain and Group Uncertain, and the test was run 10 000 times (the scores were re-shuffled prior to each test). The proportion of test statistic values that were higher than the observed value ($W = 816$) was $p = .008$.

Following up this analysis, Wilcoxon signed-rank tests performed for each group revealed the proportion of correct responses for the predictive and the non-predictive trials was significantly different in Group Certain, $V = 134$, $p = .009$, but not in Group Uncertain, $V = 318$, $p = .506$. We again ran permutation tests for each group to approximate the likelihood of observing these test statistic values by chance. The values for the proportion of correct responses were randomly shuffled between the predictive and the non-predictive trials for each participant individually, and we ran the test comparing the two trial types in each group 10 000 times. For Group Certain, the proportion of samples that were lower than the observed value ($V = 134$) was $p = .003$, while the proportion of samples lower than the value observed in Group Uncertain ($V = 318$) was $p = .755$.

There has been some interest, recently, in reporting the reliability of tests in both cognitive psychology in general (e.g., Hedge et al., 2018) and, more specifically, in the study of learning and attentional capture (e.g., Garre-Frutos, et al., 2024). Following this, we were keen to investigate the reliability of our observation that Group Certain had higher recognition accuracy for the predictive relative to the non-predictive test trials; see Appendix 5 for the details of this analysis. We also analysed reaction times (RTs) during the recognition task to investigate whether any differences in accuracy could have been a result of a speed-accuracy trade-off (e.g., participants could have taken longer to respond on the predictive test trials, contributing to their higher accuracy on these trials). However, the RTs did not vary significantly by predictiveness or group; see Appendix 3 for the complete analysis of the RTs.

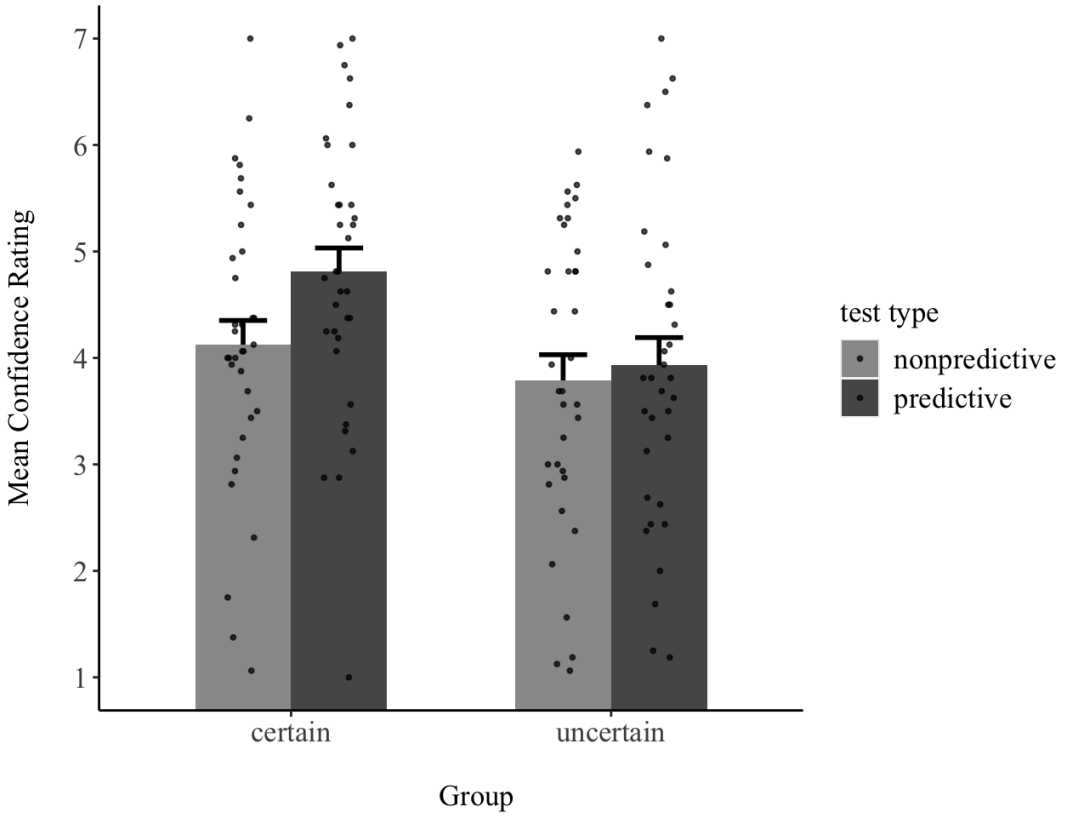
Recognition confidence ratings. During the recognition task, every stimulus selection response was followed by a confidence rating. For each participant, the confidence ratings were averaged across the predictive and the non-predictive test trials. Figure 5 shows the mean confidence ratings for the predictive and non-predictive stimuli for Groups Certain and Uncertain, which reveals that for Group Certain, confidence was higher for the predictive than the non-predictive stimuli and that, overall, there was a trend for confidence to be lower in Group Uncertain than in Group Certain. A two-way mixed ANOVA of confidence ratings with the between-subjects factor of factors of group (Certain vs. Uncertain) and a within-subjects factor predictiveness (predictive & non-predictive test trials) supported these observations. There was a main effect of predictiveness, $F(1, 68) = 9.88$, $p = .003$, $\eta_p^2 = .13$, but the main effect of group just failed to reach significance, $F(1, 68) = 3.91$, $p = .052$, $\eta_p^2 = .05$. There was, however, a significant interaction between uncertainty and predictiveness, $F(1, 68) = 4.22$, $p = .044$, $\eta_p^2 = .06$. Participants gave significantly higher certainty ratings on the predictive relative to the non-predictive test trials in Group Certain, $F(1, 34) = 10.53$, $p = .003$, $\eta_p^2 = .24$, but not in Group Uncertain, $F(1, 34) = 0.81$, $p = .376$, $\eta_p^2 = .02$. A Bayesian

mixed ANOVA with the same factors showed anecdotal support for the effect of group, $BF_{Inclusion} = 1.51$, but moderate support for the effect of predictiveness, $BF_{Inclusion} = 9.66$. There was anecdotal support for the interaction between group and predictiveness, $BF_{Inclusion} = 1.57$. In Group Certain, there was strong support for the difference in the mean ratings between the predictive and the non-predictive trials, $BF_{Inclusion} = 13.33$, but anecdotal support for the null in Group Uncertain, $BF_{Inclusion} = 0.34$.

We also investigated the simple effect of group at each level of predictiveness. On the predictive test trials, participants provided significantly higher confidence ratings in Group Certain relative to Group Uncertain, $F(1, 68) = 6.68$, $p = .012$, $\eta_p^2 = 0.09$; however, the effect of group was not significant on the nonpredictive test trials, $F(1, 68) = 1.04$, $p = .311$, $\eta_p^2 = 0.02$. Two Bayesian ANOVAs showed consistent results: there was moderate support for the effect of group on the predictive test trials, $BF_{Inclusion} = 4$, but anecdotal support for the null on the non-predictive test trials, $BF_{Inclusion} = 0.38$.

Figure 5

Recognition Test Confidence Ratings in Experiment 1: Mean Confidence Ratings to the Predictive and the Non-predictive Test Trials in Groups Certain and Uncertain



Note. Bars represent mean ratings; points represent confidence ratings for individual participants; error bars represent +/-1 SE.

Discussion

In Experiment 1, participants were presented with different compounds of stimuli and asked to predict which of two outcomes would follow. On every trial, one pair of stimuli was predictive of the outcome that followed, while the other was non-predictive. For participants in Group Certain, the relationships between the predictive pairs of stimuli and the outcomes were deterministic; for Group Uncertain, however, these relationships were probabilistic. By the end of training, both groups had learned the relationships between the compounds of

stimuli and the outcomes at an above chance level, and similar to Beesley et al. (2015), accuracy was higher for Group Certain relative to Group Uncertain. During the test phase, participants completed a recognition task, the purpose of which was to determine the accuracy of the within-compound associations that formed between the pairs of predictive and non-predictive stimuli within each compound. The results showed that Group Certain had higher recognition accuracy for the predictive relative to the non-predictive stimuli, but this difference was not observed in Group Uncertain. Furthermore, there was no overall difference in accuracy between Groups Certain and Uncertain. When participants were asked to rate how confident they were in their selections during the recognition task, their confidence ratings showed a similar pattern to that observed for accuracy: participants in Group Certain, but not in Group Uncertain, provided higher confidence ratings for the predictive relative to the non-predictive stimuli.

The hypothesis of Experiment 1 was that the representation of stimuli which possess high predictive validity or outcome uncertainty would be more accurate than stimuli which were non-predictive or followed by a deterministic outcome. This followed from (a) attention and associability being higher to stimuli which are predictive of an outcome, or followed by surprising outcomes (e.g., Esber & Haselgrove, 2011; Le Pelley et al. 2016; Pearce & Mackintosh, 2010) and (b) the representation of stimuli being more accurate when they possess stronger associations among the elements that constitute them (e.g., Hebb, 1949; McLaren & Mackintosh, 2000). The results of Experiment 1, using measures of accuracy and confidence, were consistent with the idea that the representations of predictive pairs of stimuli were more accurate than non-predictive pairs of stimuli. It is worth emphasising that in all conditions of the current experiment the relationship between the elements that comprised the predictive (AB and CD) and non-predictive (WX and YZ) components of the compounds was the same. That is to say, A was paired with B, and C was paired with D, just

as often as W was paired with X, and Y was paired with Z. Thus, any contribution of mere exposure to the formation of within-compound associations, and hence an accurate representation, was equated. An interpretation of the greater accuracy for within-compound association for the predictive stimuli is that these stimuli acquired more attention, and hence associability, than the non-predictive stimuli (e.g., Mackintosh, 1975).

However, the current experiment found no evidence for better within-compound associations in Group Uncertain relative to Group Certain. It is possible that our manipulation of uncertainty was not sufficiently powerful to detect an advantage here. Other studies that have examined the impact of outcome uncertainty on associability have employed more substantial levels of uncertainty. Kattner (2015), Livesey et al. (2011) and Chao et al. (2024) all trained participants with a 50/50 distribution between O1 and O2 presentations on uncertain trials, and Beesley et al. (2015) employed a 70/30 distribution. Perhaps if our manipulation of uncertainty had been stronger than an 80/20 distribution then we would have observed superior within-compound learning in Group Uncertain. There are at least two reasons for discrediting this possibility. First, the manipulation of uncertainty in the current experiment was not without effect: throughout training, participants' accuracy was consistently higher in Group Certain than in Group Uncertain, even when the measure of performance (proportion of probable outcome responses) could achieve 100% in both groups. Second, during the recognition test, there was a trend for confidence ratings to be lower in Group Uncertain. Thus, if anything, the effect of outcome uncertainty was to attenuate the formation of the within-compound associations, not the opposite. This pattern of results is consistent with the results from Beesley et al. (2015), who showed that predictive stimuli were more associable than non-predictive stimuli following deterministic training, but not probabilistic training, and that there was no overall associability advantage following probabilistic training.

It is worthwhile considering why, for the uncertain group, accuracy on the predictive test trials was no higher than on the nonpredictive test trials. One possibility is that this effect is in some way correlated with learning during training. For example, it may be that participants who struggled to distinguish the predictive and non-predictive stimuli during training (and thus showed poorer learning) subsequently showed no difference in within-compound associations between these two types of stimuli. To examine this possibility, we re-analysed the test data from Experiment 1, with participants who achieved an accuracy of at least .60 across training (see Appendix 4). These results revealed that, when analysed in this fashion, the certain and uncertain groups did not differ – both the certain and uncertain groups showed differences in accuracy between the predictive and non-predictive stimuli. Thus, the effect of uncertainty on representational accuracy seems to reflect the accuracy of cue-outcome learning during stage 1, rather than something about uncertainty per se.

Thus far we have considered the difference in accuracy between the predictive and non-predictive test trials in Group Certain to be a result of direct within-compound associations that formed during the training phase. However, there is an alternative possibility that is worth considering: performance at test may have been indirectly mediated by the outcome. It is possible that during the training phase participants acquired (forward) associations from the stimuli to the outcomes, as well as (backward) associations from the outcomes to the stimuli (e.g., Arcediano et al, 2003; 2005; Honey et al., 2020; Spetch et al, 1981). If this were the case, then during the test trials with the predictive stimuli a bias should be present towards selecting, for example, the correct stimulus B over the incorrect stimulus C when stimulus A is presented as a target. This follows because, during training, in addition to a forward association from A to O1, participants could have also acquired a backward association from O1 to B, resulting in an A-O1-B associative chain. Consequently, a spread of associative strength is possible from A to B that is mediated indirectly via O1 rather than

directly from A to B. There will be no such indirect mediation from A to C, however, because C was never paired with O1 during training. Applying the same analysis to the tests with the non-predictive stimuli results in there being no systematic bias in, for example, selecting the correct stimulus X over the incorrect stimulus Y when stimulus W is presented as a target. This follows because the non-predictive stimuli (W, X, Y and Z) by definition all have equivalent forward and backward associations with O1 and O2. Consequently, presenting W will result in an equivalent spread of associative strength from W through O1 and O2 to X and Y.

The mediated outcome account suggests that recognition accuracy was higher to the predictive relative to the non-predictive stimuli because two of the stimuli presented at test were indirectly linked with a common outcome, rather than because two of the stimuli became more directly linked to each other. The aim of Experiment 2 was to test this possibility.

Experiment 2

The purpose of Experiment 2 was to evaluate the outcome-mediation analysis that was developed for the results of Group Certain from Experiment 1. The crux of this analysis rests on the idea that, during test trials with the predictive stimuli in which a target stimulus (e.g., A) was presented and participants were given a choice between a correct (e.g., B) and an incorrect (e.g., C) option, the correct option was more likely to be selected because it shared an outcome in common with the target (e.g., O1), which the incorrect option did not. This bias is a consequence of the design of Experiment 1 because, for the predictive stimuli, only two stimuli were paired with each outcome (i.e., A and B were only paired with O1; C and D were only paired with O2). Thus, in order to generate an alternative, incorrect, option to accompany the correct stimulus at test, these two options would necessarily be paired with

different outcomes during training. Given that the difference in performance between the predictive and the non-predictive stimuli was only observed in Group Certain in Experiment 1, in Experiment 2 we expanded upon the design of this group to overcome this issue. Thus, in addition to the four pairs of stimuli used in Experiment 1 (predictive: AB and CD; non-predictive: WX and YZ), an additional four pairs of stimuli were employed. Now, stimulus pairs EF and GH were also predictive of O1 and O2 respectively, and pairs PQ and RS were also non-predictive. The eight different pairs of stimuli were arranged into 16 different compounds so that every predictive pair was accompanied with every non-predictive pair (see Table 2).

Table 2

Design of Experiment 2

Training phase			**Test phase**		
Compound	Outcome Probability		Trials		
	O1	O2	Predictive Same Outcome	Predictive Different Outcome	Non-predictive
ABWX	1	0	A: **B** vs F		W: **X** vs Z
ABYZ	1	0		A: **B** vs D	W: S vs **X**
ABPQ	1	0	B: E vs **A**		X: **W** vs Y
ABRS	1	0		B: C vs **A**	X: R vs **W**
CDWX	0	1	C: **D** vs H		Y: **Z** vs X
CDYZ	0	1		C: **D** vs B	Y: Q vs **Z**
CDPQ	0	1	D: G vs **C**		Z: **Y** vs W
CDRS	0	1		D: A vs **C**	Z: P vs **Y**
EFWX	1	0	E: **F** vs B		P: **Q** vs Z
EFYZ	1	0		E: **F** vs H	P: S vs **Q**
EFPQ	1	0	F: A vs **E**		Q: **P** vs Y
EFRS	1	0		F: G vs **E**	Q: R vs **P**
GHWX	0	1	G: **H** vs D		R: **S** vs X
GHYZ	0	1		G: **H** vs F	R: Q vs **S**
GHPQ	0	1	H: C vs **G**		S: **R** vs W
GHRS	0	1		H: E vs **G**	S: P vs **R**

Note. Bold typeface indicates the correct option on each test trial.

The benefit of this expanded training design is that it permits us to generate recognition test trials that involve correct and incorrect options that were paired with the same outcome during training. The column entitled “Predictive Same Outcome” in Table 2 shows the test trials in which both the correct and incorrect stimulus options were paired with the same outcome as the target stimulus during training. For example, the test trial in which A was presented as a target and B and F were presented as correct and incorrect options, all employed stimuli that were paired with O1 during training. Like Experiment 1, the

recognition test also included test trials in which the correct and incorrect options were paired with different outcomes during training. The column entitled "Predictive Different Outcome" in Table 2 shows these test trials. For example, the test trial in which A was presented as a target and B and D were presented as correct and incorrect options employed a target and a correct option that were paired with O1 during training and an incorrect option that was paired with O2. Finally, like Experiment 1, recognition tests were included with the non-predictive stimuli which, by definition, were all paired equally frequently with O1 and O2.

On the basis of the results of Experiment 1, we anticipate seeing better recognition performance to the predictive different-outcome test trials than the non-predictive test trials. The key question for Experiment 2 is how will performance on the predictive same-outcome test trials relate to this? One possibility is that the better performance on predictive vs non-predictive test trials is driven by superior within-compound associations alone, in which case the identity of the associated outcomes of the test stimuli should not matter during the test, and performance during the predictive same-outcome and predictive different-outcome test trials should be equivalent. Alternatively, if outcome mediation was partially responsible for the difference between the predictive and the non-predictive trials, then the performance on the predictive same-outcome trials should be lower to that on the predictive different-outcome trials (since participants cannot rely on outcome mediation), but still higher than that on the non-predictive trials (if within-compound associations were partially responsible for the difference). Finally, if outcome mediation was entirely responsible for the effect of predictive validity, performance on the same-outcome trials should be comparable to that on the non-predictive trials.

Method

Participants.

One hundred forty-seven participants from the UK, USA and Australia were recruited in the same way as for Experiment 1. Participants were again excluded if they did not complete the experiment, or if they made more than one attempt at beginning the training stage. There were 117 participants following the exclusions: 40 men, 31 women, and 46 participants who did not report their gender. There were 43 participants who did not report their age; the age range of the remaining participants was 18-30 years (M = 25.8, SD = 3). A larger sample size was recruited in Experiment 2 (n = 117 vs n = 35 in Group Certain from Experiment 1) in order to permit exclusions based on poor learning in phase 1 (see: results section, Experiment 2). The experiment took approximately 50 minutes to complete, and participants were given a £10 inconvenience allowance for their participation.

Stimuli and Apparatus.

The stimuli involved the same eight images of cluttered bedrooms that were used in Experiment 1, as well as eight additional images of cluttered bedrooms generated by DALL·E 1 (see Appendix 2).

Design and Procedure.

Training phase. The instructions, timing of events, stimulus locations, and stimulus sizes in Experiment 2 were identical to Experiment 1. The training phase involved 16 different trial types: each of the four predictive pairs of stimuli (AB, CD, EF, GH) was presented in compound with each of the four non-predictive pairs (WX, YZ, PQ, RS). See Table 2 for a list of the training trial types. As was the case in Experiment 1, there were 20 blocks of training trials, but for Experiment 2, each block comprised sixteen, rather than eight

trials - one trial for each trial type. Therefore, there were 320 training trials in total. The remainder of the procedure was as described in Experiment 1: the images were randomly assigned to the different stimuli at the start of experiment for each participant and participants were presented with four images on every trial and asked to assign these to the correct pressure group (O1/O2). The positions of the four stimuli were randomly determined on every trial in the same manner as for Experiment 1.

Test Phase. The recognition test followed the same procedure as in Experiment 1. The recognition test involved 32 test trials (see Table 2). Sixteen test trials comprised non-predictive stimuli and 16 trials comprised predictive stimuli. Of the test trials that comprised predictive stimuli, half employed an incorrect option that was paired with the same outcome as the target during training (predictive same-outcome trials), and half employed an incorrect option that was paired with a different outcome as the target during training (predictive different-outcome trials). Each predictive and non-predictive stimulus served equally frequently as a target stimulus (twice), and each predictive and non-predictive stimulus served equally frequently as correct and incorrect choices (again, twice each). The position of the correct and incorrect options was randomised on every trial, and the order of trial presentation was randomised across the test.

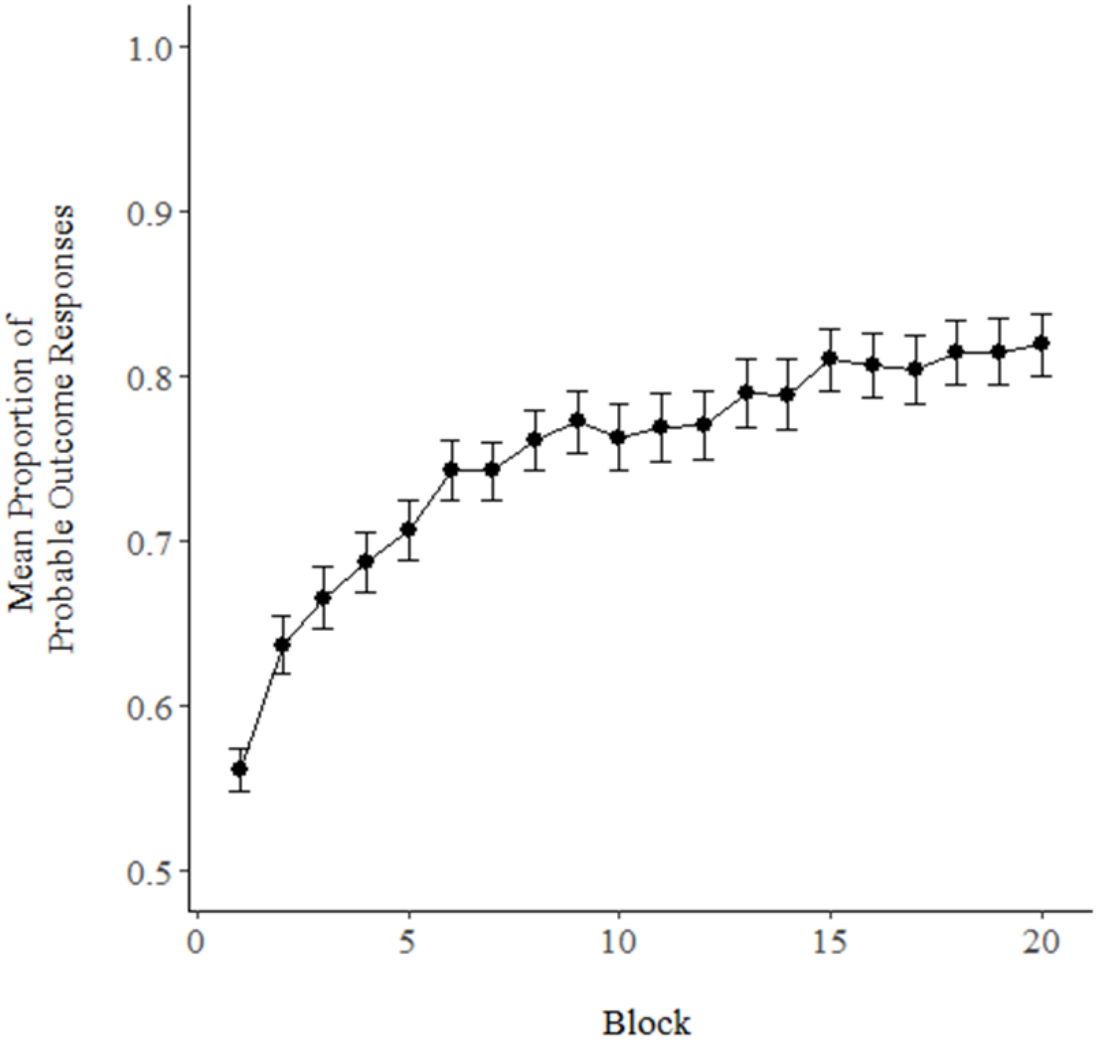
Results

Training phase accuracy. In keeping with the analysis of the training phase of Experiment 1, for each participant we calculated the proportion of correct responses for each block. Figure 6 shows the mean proportion of correct responses across the training blocks. There was a significant increase in accuracy across the training blocks; a one-way ANOVA of individual proportion of correct responses with the within subjects factor of block revealed

significant effect of block, $F(6.68, 775.33) = 39.17$, $p < .001$, $\eta_p^2 = .25$. Accuracy was above chance (0.5) during the last block of training, $t(116) = 16.66$, $p < .001$, $d = 1.54$.

Figure 6

Training Accuracy in Experiment 2: the Mean Proportion of Correct Responses across the Training Phase



Note. Error bars represent +/-1 SE.

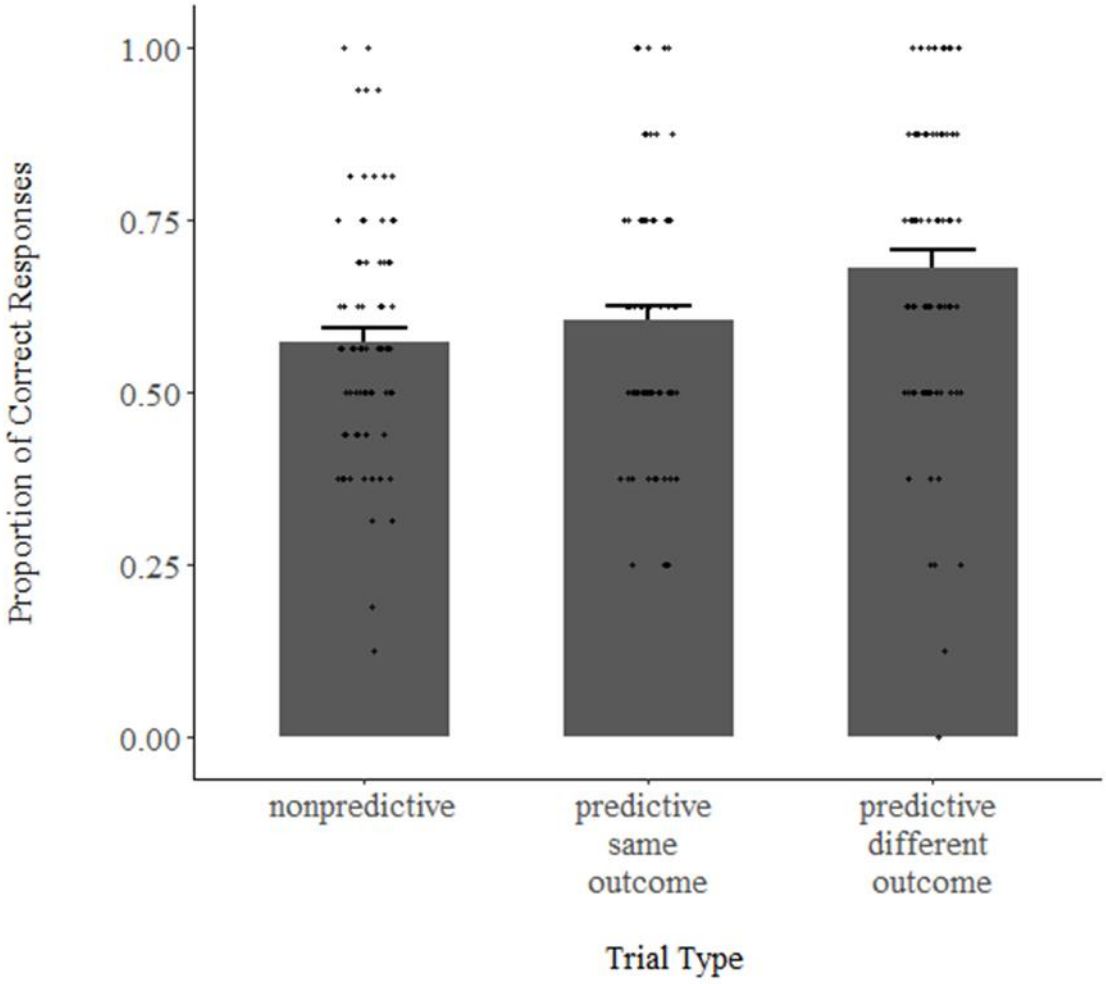
Recognition test accuracy. An effect of predictive validity is only to be expected if participants used the feedback in phase 1 to learn the correct stimulus-outcome relationships.

Therefore, we adopted the criterion used elsewhere (e.g., Le Pelley et al., 2011; Le Pelley & McLaren, 2003) in which participants were excluded from all subsequent analyses if they had a mean proportion of correct responses during the training phase below 0.6 (final n = 85). A similar exclusion criterion was not appropriate for Experiment 1 as it would selectively impact the sample for Group Uncertain.

For each participant, the proportion of correct stimulus selections was calculated for each of the three trial types: predictive same outcome, predictive different outcome, and non-predictive. Figure 7 shows the proportion of correct stimulus selections for each trial type. Overall, correct responding was higher on the predictive different outcome trials relative to both predictive same outcome and non-predictive trials, while the accuracy was comparable for the latter two trial types.

Figure 7

Recognition Test Accuracy in Experiment 2: The Mean Proportion of Correct Stimulus Selections by Trial Type



Note. Bars show mean proportion correct responses; points show proportion correct responses for individual participants; error bars represent +/-1 SE.

A one-way repeated-measures ANOVA of correct stimulus selections with a within-subjects factor of trial type was conducted. This analysis revealed a significant main effect of trial type, $F(2, 168) = 9.41$, $p < .001$, $\eta_p^2 = 0.1$, with a corresponding Bayesian analysis showing substantial support for the main effect of trial type, $BF_{Inclusion} = 163.52$. Pairwise comparisons revealed that the accuracy on the predictive different outcome trials was significantly higher than on the non-predictive trials, $t(84) = 4.10$, $p < .001$, $d = 0.44$, $BF_{10} = 204.07$, replicating the effect observed in Group Certain in Experiment 1. The accuracy on the predictive same outcome trials was significantly lower than on the predictive different outcome trials, $t(84) = 2.94$, $p = .004$, $d = 0.32$, $BF_{10} = 6.42$, while accuracy did not differ significantly between the nonpredictive trials and the predictive same outcome trials, $t(84) = 1.23$, $p = .222$, $d = 0.13$, $BF_{10} = 0.25$. The proportion of correct stimulus selections was above chance (0.5) for all three trial types, $ts(84) > 3.83$, $ps < .001$, $ds > 0.42$.

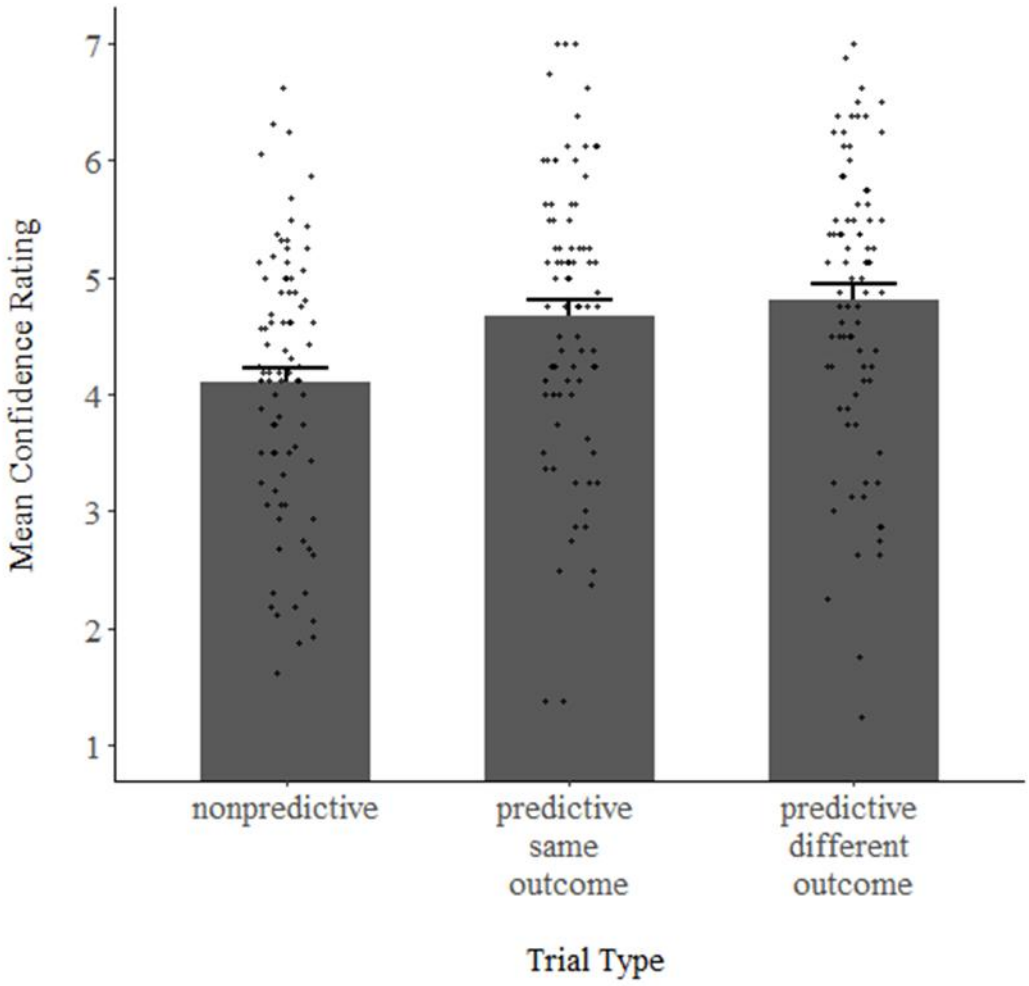
We again examined the reliability of the effect of predictiveness in the same way as in Experiment 1; the details of this analysis are reported in Appendix 5. As was the case in Experiment 1, we investigated whether the differences in recognition accuracy between the three trial types were accompanied by any differences in the RTs. The analysis suggests that participants had shorter RTs on both types of the predictive test trials relative to the non-predictive test trials; see Appendix 3 for the full analysis of the RTs.

A Shapiro-Wilk test indicated that the proportions of correct responses were not normally distributed across participants, $W = 0.97$, $p < .001$. Therefore, we conducted Friedman's ANOVA as a non-parametric equivalent to the repeated-measures ANOVA. This test indicated that there was a significant difference between the three trial types, $\chi^2(2) = 18.10$, $p < .001$. Post-hoc tests for Friedman's ANOVA revealed that the difference between the nonpredictive and the predictive same outcome trials was not significant (observed = 18.5, critical = 31.21). The difference between the nonpredictive and the predictive different outcome trials was significant (observed = 52, critical = 31.21), as was the difference between the predictive same outcome and the predictive different outcome trials (observed = 33.5, critical = 31.21).

Recognition confidence ratings. To investigate whether the confidence ratings varied with trial type, for each participant, we calculated the mean rating for the non-predictive, predictive same outcome, and predictive different outcome trials. Figure 8 shows the mean confidence rating by trial type. Confidence ratings were higher, overall, following responses on the predictive relative to the non-predictive trials – again reproducing the result observed in Group Certain from Experiment 1. A one-way frequentist and Bayesian ANOVAs of individual confidence ratings revealed a significant effect of trial type, $F(1.67, 140.7) = 33.49$, $p < .001$, $\eta_p^2 = .29$, $BF_{Inclusion} > 10{,}000$.

Figure 8

Recognition Test Confidence Ratings for Experiment 2: Mean Confidence Ratings by Trial Type



Note. Bars show grand mean ratings; points show mean confidence ratings for individual participants; error bars represent +/-1 SE.

In contrast to the patterns observed in recognition accuracy, the mean ratings on the non-predictive trials were lower relative to both the predictive different outcome trials, $t(84) = 6.78$, $p < .001$, $d = 0.74$, $BF_{10} > 10{,}000$, and the predictive same outcome trials, $t(84) = 5.86$, $p < .001$, $d = 0.64$, $BF_{10} > 10{,}000$. The mean ratings did not significantly differ between

the predictive same outcome and the predictive different outcome trials, $t(84) = 1.86$, $p = .066$, $d = 0.20$, $BF_{10} = 0.62$.

Discussion

In Experiment 2, participants were required to predict which of two outcomes followed different compounds that were composed of two pairs of stimuli. On separate trials, stimulus pairs AB and EF were predictive of O1, and pairs CD and GH were predictive of O2. Other pairs of stimuli (WX, YZ, PQ or RS) were non-predictive. In keeping with Group Certain from Experiment 1, the relationship between the predictive pairs of stimuli and the outcomes was deterministic. The inclusion of predictive stimulus pairs EF and GH allowed us to incorporate "predictive same outcome" trials at test, which included correct and incorrect response options that had the same associated outcome as the target. If recognition accuracy for these predictive trials were higher than those of non-predictive cues, this would indicate strong evidence for within-compound learning that is not mediated by outcome association learning.

The results of Experiment 2 reproduced the effect observed in Group Certain from Experiment 1: participants showed higher recognition memory on the predictive different-outcome trials relative to the non-predictive trials.. Interestingly, performance on the predictive same-outcome trials was lower than on the predictive different-outcome trials, suggesting that outcome mediation played a role in the effect observed in group Certain in Experiment 1. For example, when participants were required to choose between B or C as being the associate of A, their knowledge that A and B were both paired with O1, but that A and C were not, allowed them to perform more accurately on these trials relative to the non-predictive trials. Given that performance on the predictive same-outcome trials was comparable to that on the non-predictive trials ($BF_{10} = 0.25$), these patterns suggest that

outcome mediation was entirely responsible for the difference observed between the predictive different-outcome trials and the non-predictive trials.

However, recognition accuracy was above chance on all three types of test trials. Importantly, this was true for the predictive-same and the non-predictive test trials, in which the correct-option could not be selected over the incorrect-option on the basis of these options being previously paired with different outcomes. This implies that direct (i.e., non-outcome-mediated) within-compound associations did develop between the stimuli within each compound during training. However, because performance did not differ between the non-predictive and predictive-same trials suggests that these within-compound associations did not vary in strength as a function of the predictive validity of the stimuli (i.e., the direct A-B association was as strong as the direct W-X association). Together, then, the results of Experiment 2 suggest that the representational accuracy is determined by two factors: direct (i.e., within-event) associations and indirect (i.e., outcome mediated) associations.

The focus of the current studies was upon the extent to which learning influenced the representational accuracy of stimuli – i.e., the extent to which they are veridical. It is natural, then, that our principle dependent variable at test assessed the extent to which learnt behaviour was in correspondence with objective reality – recognition accuracy (Figures 4 and 7). However, in keeping with other studies of recognition memory we also collected data on participants' confidence ratings which, although often considered to tap the same memory representation, can sometimes be only weakly correlated with accuracy (e.g., Busey et al., 2000). For Experiment 2, the pattern of results observed in confidence ratings (Figure 8) did not fully correlate with the patterns of recognition accuracy: confidence ratings did not differ between the two types of predictive test trials, whereas recognition accuracy did. These results suggest that the confidence ratings were modulated by a variable other than the direct

within-compound associations that constituted the structure of the stimuli. It is possible that, similarly to their selections, participants' confidence ratings were based on their knowledge of the associated outcomes. For example, participants may have provided high confidence ratings whenever they were able to choose an option that shared the same outcome as the target, which would have been the case on both types of the predictive test trials. In contrast, participants could not rely on their knowledge of the associated outcomes on the non-predictive trials, resulting in lower confidence ratings.

General Discussion

In two experiments participants were presented with compounds of four relatively complex stimuli and asked to learn the predictive relationships between these compounds and one of two outcomes. Within these compounds, pairs of stimuli were consistently presented together (e.g., A and B were always presented as pairs, as were W and X). In each experiment, some pairs were established as predictive of the trial outcome, and other pairs of stimuli were established as non-predictive. In Experiments 1 and 2 the relationship between the compounds and the outcomes was deterministic: for example, O1 always followed compound ABWX. In Experiment 1, however, we also included a condition in which the compound-outcome relationship was probabilistic: for example, O1 followed compound ABWX only 80% of the time. Participants learned the relationships between compounds and outcomes, but, unsurprisingly, learning was attenuated when outcome uncertainty was introduced. The crucial findings of the current experiments came from the subsequent test phase in which participants were presented with one stimulus from each pair and asked to select the stimulus that accompanied it during training. Participants were more accurate in making this selection (and were more confident about their decision) when the pairs of stimuli had been predictive of an outcome during training than when the pairs had been non-predictive (Experiment 1 and 2). However, this effect was not apparent when the compound-

outcome relationship during training was probabilistic in Experiment 1 (but see Appendix 4 for a further analysis). Furthermore, overall recognition accuracy was equivalent when training was either deterministic or probabilistic.

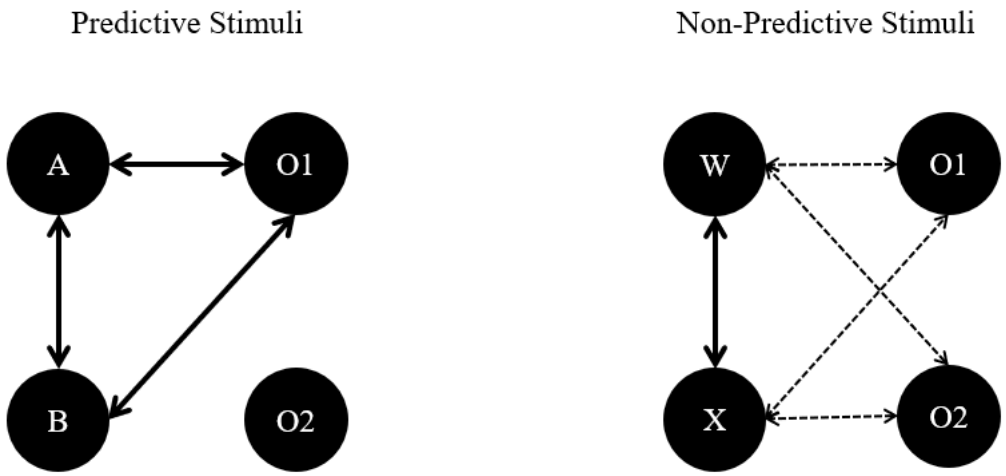
Experiment 2 provided information about the associative structure of the apparent within-event learning that was evident in the Certain group from Experiment 1. When test trials were given in which participants were required to select the associate of a target from two options - both of which were paired with the same outcome during training - the effect of predictive validity was abolished. This result implies the operation of an indirect association between stimuli that was mediated through a representation of the outcome they were paired with during training. Finally, the observation, in Experiment 2, of within-compound associative knowledge that was above chance on both the non-predictive and predictive-same test trials implies a role for the acquisition of direct-within compound associations during training. However, the absence of any effect of predictive validity on these trials suggests that these associations were a consequence of mere exposure.

Experiment 2 revealed the presence of both direct and indirect associations among the stimuli during test, and Figure 9 shows a simplified sketch of the associations that may have formed among pairs of predictive and non-predictive stimuli, and their outcomes, on an ABWX trial. Because A and B are perfectly correlated with the presentation of O1 (tables 1 and 2) we can assume a strong excitatory association formed between stimuli A and B, and also between A and O1 and O1 and B. Consequently, presenting stimulus A will activate stimulus B through two relatively strong routes: one direct (A-B), and the other indirect (A-O1-B). In contrast, although the presentations of W and X are perfectly correlated with each other, W and X are entirely un-correlated with O1 and O2, thus the indirect excitatory associations between W and X via O1 and O2 will be comparatively weak. It is thus likely

that the sum of the direct and indirect associations will be greater for the predictive than the non-predictive stimuli, and recognition accuracy should be higher, which is the result observed in in the Certain group in Experiment 1 (and Experiment 2). Although this effect was attenuated when outcome uncertainty was introduced (Uncertain Group - Experiment 1), this difference between Group Certain and Group Uncertain appears to be due to the higher number of participants in Group Uncertain who did not learn the distinction between the predictive and the non-predictive cues (see Appendix 4). These participants may not have formed forward (and/or backward) associations between the predictive cues and the outcomes that were strong enough to support outcome mediation at test.

Figure 9

The structure of within- and between-event associations for two example pairs of predictive (L) and non-predictive (R) stimuli



Note. Letters refer to predictive (A and B) and non-predictive (W and X) pairs of stimuli. O1 and O2 refer to Outcome 1 and Outcome 2. Bold arrows refer to strong excitatory associations, dashed lines refer to weaker excitatory associations.

The motivation behind our studies was based on two premises: (a) the accuracy of a stimulus representation should depend on the associability of its elements, and (b) if elements of a stimulus compound were established with high predictive validity (e.g., Mackintosh, 1975) or outcome uncertainty (Pearce & Hall, 1980), then those elements should become better associated (cf, McLaren & Mackintosh, 2000). However, our results found no evidence that predictive validity or outcome uncertainty resulted in more accurate within-compound associations between cues presented together. Instead, participants responded as if they formed (a) direct within-compound associations as a consequence of mere exposure, plus (b) indirect associations between cues that shared the same outcome irrespective of whether they had in reality been paired. Thus, these results comprise a paradox: while previous studies have demonstrated elevated attention towards stimuli with high predictive validity and outcome uncertainty, our study suggests that such elevated attention towards these stimuli might not result in more accurate within-compound associations. This is particularly perverse in the face of the suggestion that a proposed effect of outcome uncertainty is to initiate a state of exploratory attention within compounds of complex stimuli (e.g., Beesley et al., 2015; Luque et al., 2017).

The outcome mediation account, suggested as a possible explanation for our results, makes no appeal to the acquisition of differential amounts of attention paid to stimuli as a function of learning, and could provide a partial resolution to the paradox introduced in the previous paragraph. While the motivation for the current studies was to determine if predictive validity and outcome uncertainty influence the acquisition of within-compound associations, it is possible that the opposite may in fact be true: the strength of within-compound associations influences the associability of a stimulus. We have already discussed how theories such as McLaren and Mackintosh (2000) use the acquisition of within-compound associations to reduce the trial-to-trial variability in the spread of activation that

comes from sampling a complex stimulus. One might reasonably expect that a consequence of this reduction in variability could also be a corresponding increase in stimulus associability: learning about a stimulus should progress more rapidly if the same set of its elements is processed on each trial, compared to the case in which different subsets of elements are sampled and processed on each trial (see also: Byrom & Murphy, 2016; Harris, 2006). Similarly, we might also expect stimuli with stronger (or more) within-compound associations to attract longer dwell times. This follows because the probability of an attentional response shifting from one element to another element within the same stimulus (rather than shifting to an element of a different stimulus) will be a function of learning about the structure of that stimulus (e.g., Arato et al., 2024) - i.e., the strength of its within-compound associations. Other things being equal, then, we might expect to see longer dwell times to predictive than to non-predictive stimuli. Together, then, the acquisition of stronger within-compound associations in predictive than non-predictive stimuli would provide an account for the associability of these stimuli that aligns with the theory proposed by Mackintosh (1975), as well as the studies that have shown elevated attention towards predictive stimuli. An obvious caveat of this analysis, however, is to also explain why overt attention is often enhanced to stimuli associated with outcome uncertainty (e.g., Beesley et al., 2015).

A computational model proposed by Honey et al. (2020), HeiDI, is in line with the outcome mediation account of our results. In contrast to standard models of Pavlovian conditioning (e.g., Rescorla & Wagner, 1972; Mackintosh, 1975; Pearce & Hall, 1980), Honey et al. emphasise the importance of reciprocal forward and backward associations being formed between stimuli and outcomes during conditioning (cf. Kahana, 2002). Such reciprocal associations prove useful as an account of higher-order conditioning phenomena, such as backward sensory preconditioning (Ward-Robinson and Hall, 1996; for a detailed

explanation, see Honey et al., 2022), that are difficult to reconcile with standard accounts of Pavlovian conditioning. The model also specifies learning rules for the kind of representation-mediated learning that we propose occurred via the outcomes in our experiments. The model achieves representation-mediated learning by suggesting that the salience of the associatively activated element (i.e., the outcomes at test in our experiments) is proportional to the associative strength it shares with the presented stimuli. This suggestion is in line with the difference that we see between Group Certain and Group Uncertain in Experiment 1. When our analyses excluded participants who performed close to chance during the training phase (i.e., participants who, we assume, had weaker cue-outcome associations), the difference in test accuracy between the predictive and the non-predictive cues was similar for Group Certain and Group Uncertain. This is to be expected if associative activation of the outcomes at test is proportional to the reciprocal associations that form during the training phase.

Our representation of stimuli within the environment is far from veridical (e.g. Simons and Chabris, 1999). However, experience can improve this situation. At face value it appears that task relevancy (predictive validity) selectively improves the accuracy of the representation of a predictive stimulus relative to the representation of a non-predictive stimulus. We suggest that this apparent effect is driven by two associative pathways: (a) a direct (within-event) association between simultaneously presented stimuli and (b) an indirect (between-event) association between the same stimuli that is mediated by the activation of their common outcome. Together, these pathways have the potential to explain some commonly observed properties of predictive stimuli (such as their relatively superior associability, and overt attention). A challenge that remains is to better understand the circumstances under which outcome uncertainty might also have an impact upon representational accuracy.

References

Arato, J., Rothkopf, C. A., & Fiser, J. (2024). Eye movements reflect active statistical learning. *Journal of Vision*, *24*(5), 1-18.

Arcediano, F., Escobar, M., & Miller, R. R. (2003). Temporal integration and temporal backward associations in human and nonhuman subjects. *Animal Learning & Behavior*, *31*(3), 242-256.

Arcediano, F., Escobar, M., & Miller, R. R. (2005). Bidirectional associations in humans and rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(3), 301-318.

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E. J. (2015). An introduction to Bayesian hypothesis testing for management research. *Journal of Management*, *41*(2), 521-543.

Atkinson, R. C., & Estes, W. K. (1962). Stimulus sampling theory. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. 2, pp121-268). New York: Wiley.

Ayres, J. J., Moore, J. W., & Vigorito, M. (1984). Hall and Pearce negative transfer: Assessments in conditioned suppression and nictitating membrane conditioning experiments. *Animal Learning & Behavior*, *12*(4), 428-438.

Beesley, T., Nguyen, K. P., Pearson, D., & Le Pelley, M. E. (2015). Uncertainty and predictiveness determine attention to cues during human associative learning. *Quarterly Journal of Experimental Psychology*, *68*(11), 2175-2199.

Beesley, T., & Shanks, D. R. (2012). Investigating cue competition in contextual cuing of visual search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 709–725. https://doi.org/10.1037/a0024885

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic bulletin & review*, *7*, 26-48.

Byrom, N. C., & Murphy, R. A. (2014). Sampling capacity underlies individual differences in human associative learning. *Journal of Experimental Psychology: Animal Learning and Cognition*, *40*(2), 133-143.

Cheatle, M. D., & Rudy, J. W. (1978). Analysis of second-order odor-aversion conditioning in neonatal rats: implications for Kamin's blocking effect. *Journal of Experimental Psychology: Animal Behavior Processes*, *4*(3), 237-249.

Chao, C. M., McGregor, A., & Sanderson, D. J. (2021). Uncertainty and predictiveness modulate attention in human predictive learning. *Journal of Experimental Psychology: General*, *150*(6), 1177-12002.

Cho, S. A., & Cho, Y. S. (2021). Uncertainty modulates value-driven attentional capture. *Attention, Perception, & Psychophysics*, *83*, 142-155.

Collins, L., & Pearce, J. M. (1985). Predictive accuracy and the effects of partial reinforcement on serial autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*(4), 548-564.

Collins, L., Young, D. B., Davies, K., & Pearce, J. M. (1983). The influence of partial reinforcement on serial autoshaping with pigeons. *The Quarterly Journal of Experimental Psychology*, *35*(4), 275-290.

Crick, F. (1984). Function of the thalamic reticular complex: the searchlight hypothesis. *Proceedings of the National Academy of Sciences*, *81*(14), 4586-4590.

Dobbins, I. G., Kroll, N. E. A., & Liu, Q. (1998). Confidence-accuracy inversions in scene recognition: A remember-know analysis. Journal of Experimental Psychology: Learning. Memory. & Cognition, 24, 1306-1315.

Durlach, P. J., & Mackintosh, N. J. (1986). Transfer of serial reversal learning in the pigeon. *The Quarterly Journal of Experimental Psychology*, *38*(1), 81-95.

Durlach, P. J., & Rescorla, R. A. (1980). Potentiation rather than overshadowing in flavor-aversion learning: an analysis in terms of within-compound associations. *Journal of Experimental Psychology: Animal Behavior Processes*, *6*(2), 175-187.

Easdale, L. C., Le Pelley, M. E., & Beesley, T. (2019). The onset of uncertainty facilitates the learning of new associations by increasing attention to cues. *Quarterly journal of experimental psychology*, *72*(2), 193-208.

Esber, G. R., & Haselgrove, M. (2011). Reconciling the influence of predictiveness and uncertainty on stimulus salience: a model of attention in associative learning. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1718), 2553-2561.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological review*, *57*(2), 94.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, 41(4), 1149-1160.

Garre-Frutos, F., Vadillo, M. A., González, F., & Lupiáñez, J. (2024). On the reliability of value-modulated attentional capture: An online replication and multiverse analysis. *Behavior Research Methods*, 1-18.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment?. *Psychological review*, *62*(1), 32-41.

Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human perception and performance*, *26*(1), 86-112.

Griffiths, O., & Mitchell, C. J. (2008). Selective attention in human associative learning and recognition memory. *Journal of Experimental Psychology: General*, *137*(4), 626-648.

Hall, G. (1991). *Perceptual and Associative Learning*. Oxford: Clarendon.

Hall, G., & Pearce, J. M. (1979). Latent inhibition of a CS during CS–US pairings. *Journal of Experimental Psychology: Animal Behavior Processes*, *5*(1), 31-42.

Hall, G., & Pearce, J. M. (1982). Restoring the associability of a pre-exposed CS by a surprising event. *The Quarterly Journal of Experimental Psychology Section B*, *34*(3b), 127-140.

Harris, J. A. (2006). Elemental representations of stimuli in associative learning. *Psychological review*, *113*(3), 584-605.

Haselgrove, M., Esber, G. R., Pearce, J. M., & Jones, P. M. (2010). Two kinds of attention in Pavlovian conditioning: evidence for a hybrid model of learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(4), 456-470.

Haselgrove, M., Le Pelley, M. E., Singh, N. K., Teow, H. Q., Morris, R. W., Green, M. J., Griffiths, O., & Killcross, S. (2016). Disrupted attentional learning in high schizotypy: Evidence of aberrant salience. *British journal of psychology*, *107*(4), 601-624.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior research methods*, *50*, 1166-1186.

Hogarth, L., Dickinson, A., Austin, A., Brown, C., & Duka, T. (2008). Attention and expectation in human predictive learning: The role of uncertainty. *Quarterly Journal of Experimental Psychology*, *61*(11), 1658-1668.

Holland, P. C. (1980). CS–US interval as a determinant of the form of Pavlovian appetitive conditioned responses. *Journal of Experimental Psychology: Animal Behavior Processes*, *6*(2), 155-174.

Honey, R. C., & Dwyer, D. M. (2022). Higher-order conditioning: A critical review and computational model. *Psychological Review*, *129*(6), 1338.

Honey, R. C., Dwyer, D. M., & Iliescu, A. F. (2020). HeiDI: A model for Pavlovian learning and performance with reciprocal associations. *Psychological Review*, *127*(5), 829-852.

Honey, R. C., Schachtman, T. R., & Hall, G. (1987). Partial reinforcement in serial autoshaping: The role of attentional and associative factors. *Learning and Motivation*, *18*(3), 288-300.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological review*, *99*(3), 480-517.

Ju, J., & Cho, Y. S. (2023). The modulation of value-driven attentional capture by exploration for reward information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(2), 181-197.

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & cognition*, *30*(6), 823-840.

Kattner, F. (2015). Transfer of absolute and relative predictiveness in human contingency learning. *Learning & 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*(1), 90-109.

Keele, S. W., Cohen, A., Ivry, R., Liotti, M., & Yee, P. (1988). Tests of a temporal theory of attentional binding. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 444-452.

Kersten, A. W., Goldstone, R. L., & Schaffert, A. (1998). Two competing attentional mechanisms in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1437-1458.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of mathematical psychology*, *45*(6), 812-863.

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science*, *12*(5), 171-175.

Larkin, M. J., Aitken, M. R., & Dickinson, A. (1998). Retrospective revaluation of causal judgments under positive and negative contingencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1331.

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 Section B*, *57*(3), 193-243.

Le Pelley, M. E., Haselgrove, M., & Esber, G. R. (2012). Modelling attention in associative learning: Two processes or one?. *Learning & behavior*, *40*, 292-304.

Le Pelley, M. E., & McLaren, I. P. L. (2003). Learned associability and associative change in human causal learning. *The Quarterly Journal of Experimental Psychology: Section B*, *56*(1), 68-79.

Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., & Wills, A. J. (2016). Attention and associative learning in humans: An integrative review. *Psychological bulletin*, *142*(10), 1111-1140.

Le Pelley, M. E., Turnbull, M. N., Reimers, S. J., & Knipe, R. L. (2010). Learned predictiveness effects following single-cue training in humans. *Learning & Behavior*, *38*(2), 126-144.

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*(2), 220-229.

Le Pelley, M. E., Vadillo, M., & Luque, D. (2013). Learned predictiveness influences rapid attentional capture: evidence from the dot probe task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1888-1900.

Livesey, E. J., Harris, I. M., & Harris, J. A. (2009). Attentional changes during implicit learning: signal validity protects a target stimulus from the attentional blink. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 408-422.

Livesey, E. J., Thorwart, A., De Fina, N. L., & Harris, J. A. (2011). Comparing learned predictiveness effects within and across compound discriminations. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*(4), 446-465.

Lochmann, T., & Wills, A.J. (2003). Predictive history in an allergy prediction task. In F. Schmalhofer, R.M. Young & G. Katz (Eds.), *Proceedings of EuroCogSci 03* (pp. 217222). Mahwah, NJ: Lawrence Erlbaum Associates

Luque, D., Vadillo, M. A., Le Pelley, M. E., & Beesley, T. (2017). Prediction and uncertainty in associative learning: Examining controlled and automatic components of learned attentional biases. *Quarterly Journal of Experimental Psychology*, *70*(8), 1485-1503.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114*, 159-188

McLaren, I. P. L., & Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior*, *28*(3), 211-246.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological review*, *82*(4), 276-298.

Mackintosh, N.J. and Little, L., (1969). Intradimensional and extradimensional shift learning by pigeons. *Psychonomic Science, 14*, 5-6.

Melchers, K., Lachnit, H., & Shanks, D. (2004). Within-compound associations in retrospective revaluation and in direct learning: A challenge for comparator theory. *The Quarterly Journal of Experimental Psychology Section B*, *57*(1), 25-53.

Muñiz-Diez, C., Lagator, S., Haselgrove, M., & Beesley, T (in Prep). Effects of expected and unexpected uncertainty on cue processing.

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*(4), 841-849.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological review*, *87*(6), 532-552.

Pearce, J. M., Kaye, H., & Collins, L. (1985). A comparison of the effects of partial reinforcement schedules using a within-subject serial autoshaping procedure. *The Quarterly Journal of Experimental Psychology Section B*, *37*(4b), 379-396.

Pearce, J. M., Kaye, H., & Hall, G. (1982). Predictive accuracy and stimulus associability: Development of a model for Pavlovian learning. In M. Commons, R. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analysis of behavior* (Vol. 3, pp. 241-256).

Pearce, J. M., & Mackintosh, N. J. (2010). Two theories of attention: A review and a possible integration. *Attention and associative learning: From brain to behaviour*, 11-39.

Pearson, D., Chong, A., Chow, J. Y. L., Garner, K. G., Theeuwes, J., & Le Pelley, M. E. (2024). Uncertainty-modulated attentional capture: Outcome variance increases attentional priority. *Journal of Experimental Psychology: General, 153*(6), 1628–1643.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior research methods*, *51*, 195-203.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rescorla, R. A. (1980). Simultaneous and successive associations in sensory preconditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *6*(3), 207-216.

Rescorla, R. A. (1981). Within-signal learning in autoshaping. *Animal Learning & Behavior*, *9*(2), 245-252.

Rescorla, R. A., & Colwill, R. M. (1983). Within-compound associations in unblocking. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(4), 390-400.

Rescorla, R. A., & Cunningham, C. L. (1978). Within-compound flavor associations. *Journal of Experimental Psychology: Animal Behavior Processes*, *4*(3), 267-275.

Rescorla, R. A., & Durlach, P. J. (1981). Within-Event Learning In Pavlovian Conditioning. In N. E. Spear & R. R. Miller (Eds) Information Processing in Animals: Memory Mechanisms (pp. 81-111). Psychology Press.

Rescorla, R. A., & Freberg, L. (1978). The extinction of within-compound flavor associations. *Learning and Motivation*, *9*(4), 411-427.

Roberts, A. C., Robbins, T. W., & Everitt, B. J. (1988). The effects of intradimensional and extradimensional shifts on visual discrimination learning in humans and non-human primates. *The Quarterly Journal of Experimental Psychology*, *40*(4), 321-341.

Rodríguez, G., & Alonso, G. (2011). Reinforced stimulus preexposure effects as a function of US intensity: Implications for understanding the Hall–Pearce effect. *Learning and Motivation*, *42*(2), 193-200.

Savastano, H. I., Yin, H., Barnet, R. C., & Miller, R. R. (1998). Temporal coding in Pavlovian conditioning: Hall-Pearce negative transfer. *The Quarterly Journal of Experimental Psychology Section B*, *51*(2b), 139-153.

Schmajuk, N. A., Lam, Y. W., & Gray, J. A. (1996). Latent inhibition: a neural network approach. *Journal of Experimental Psychology: Animal Behavior Processes*, *22*(3), 321-349.

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. *The psychology of learning and motivation*, *31*, 305-349.

Shepp, B. E., & Eimas, P. D. (1964). Intradimensional and extradimensional shifts in the rat. *Journal of comparative and physiological psychology*, *57*(3), 357-361.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059-1074.

Spetch, M. L., Wilkie, D. M., & Pinel, J. P. (1981). Backward conditioning: a reevaluation of the empirical evidence. *Psychological Bulletin*, *89*(1), 163-175.

Swan, J. A., & Pearce, J. M. (1988). The orienting response as an index of stimulus associability in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*(3), 292-301.

Swartzentruber, D., & Bouton, M. E. (1992). Context sensitivity of conditioned suppression following preexposure to the conditioned stimulus. *Animal Learning & Behavior*, *20*(2), 97-103.

Thorwart, A., Livesey, E. J., Wilhelm, F., Liu, W., & Lachnit, H. (2017). Learned predictiveness and outcome predictability effects are not simply two sides of the same coin. *Journal of Experimental Psychology: Animal Learning and Cognition*, *43*(4), 341 - 365.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology Section A*, *40*(2), 201-237.

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1373), 1295-1306.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97-136.

Wagner, A. R., (1978) Expectancies and the priming of STM. IN S. H. Hulse, H. Fowler, & W. K. Honig (Eds.) *Cognitive Processes in Animal Behavior* (pp. 177-209). Hillsdale, NJ: Erlbaum.

Ward-Robinson, J., & Hall, G. (1996). Backward sensory preconditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *22*(4), 395-404.

Walker, A. R., Luque, D., Le Pelley, M. E., & Beesley, T. (2019). The role of uncertainty in attentional and choice exploration. *Psychonomic Bulletin & Review*, *26*(6), 1911-1916.

Walker, A. R., Navarro, D. J., Newell, B. R., & Beesley, T. (2022). Protection from uncertainty in the exploration/exploitation trade-off. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(4), 547–568

Wasserman, E., & Berglan, L. R. (1998). Backward blocking and recovery from overshadowing in human causal judgement: The role of within-compound associations. *The Quarterly Journal of Experimental Psychology: Section B*, *51*(2), 121-138.

Welham, A. K., & Wills, A. J. (2011). Unitization, similarity, and overt attention in categorization and exposure. *Memory & Cognition*, *39*(8), 1518-1533.

Wilson, P. N., Boumphrey, P., & Pearce, J. M. (1992). Restoration of the orienting response to a light by a change in its predictive accuracy. *The Quarterly Journal of Experimental Psychology Section B*, *44*(1b), 17-36.

Appendix 1

Full Instructions for the training phase of Experiments 1 and 2

"Welcome to spy training. We've uncovered evidence that two political pressure groups, THE LIBERTY ALLIANCE and THE PROGRESS COALITION, have begun espionage activities. Huge numbers of photos are being taken by these two pressure groups. Our team has intercepted emails that contain these photos, with each email containing four photos that show different rooms. It has become clear that these rooms had been bugged for covert surveillance. But here's the problem: we don't know which email belongs to which pressure group. That's where you come in. Your spymaster, M, believes that, with training, your brain can learn to identify the connections between the photos and the two political pressure groups.

Your mission is to become an expert at linking the photos in these emails to either the Liberty Alliance or the Progress Coalition. On every trial, you will be shown four photos. Your task is to determine whether they belong to the Liberty Alliance or the Progress Coalition. At first, you will have to guess. With training and feedback, we hope you will learn to link these photos to the correct pressure group. Good luck, your country needs you."

Full Instructions for the test phase of Experiments 1 and 2

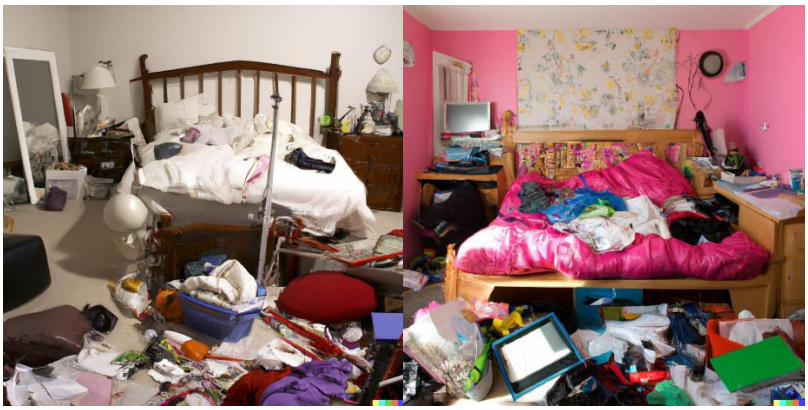
"On every trial, you will be shown a photo on the left side (see example below). There will be two other photos on the right. One of the photos on the right was shown in the same bundle as the photo on the left, whereas the other was not.

Your task is to choose the photo on the right that was shown with the photo on the left."

Appendix 2

Stimuli used to represent cues A, B, C, D and W, X, Y, Z in Experiments 1 and 2

Stimuli used to represent the additional cues E, F, G, H and P, Q, R, S in Experiment 2



Appendix 3

Analysis of Reaction Times in Experiment 1

To investigate whether any differences in recognition accuracy were accompanied by a speed-accuracy trade-off, for each participant we calculated the mean RT for stimulus selection responses separately for the predictive and the non-predictive test trials. The RTs were expressed in milliseconds and were log transformed due to a positive skew in the data. The mean RTs (and standard errors) for the predictive and non-predictive stimuli for Group Certain were 7.76 log ms (0.08) and 7.87 log ms (0.09) respectively. The mean RTs (and standard errors) for the predictive and non-predictive stimuli for Group Uncertain were 7.71 log ms (0.08) and 7.73 log ms (0.08) respectively. A two-way ANOVA of individual RTs with the factors of Group (Certain vs Uncertain) and predictiveness (predictive & non-predictive test trials) revealed no main effect of group, $F(1, 68) = 0.72$, $p = .401$, $\eta_p^2 = .01$, $BF_{inclusion} = 0.61$, a main effect of predictiveness that just failed to reach significance, $F(1, 68) = 3.88$, $p = .053$, $\eta_p^2 = .05$, $BF_{inclusion} = 0.99$, and no interaction between these factors, $F(1, 68) = 2.52$, $p = .117$, $\eta_p^2 = .04$, $BF_{inclusion} = 0.66$. There was thus no evidence for a speed-accuracy trade off. Indeed, if anything, RTs were faster, overall, to the predictive rather than the non-predictive stimuli.

Analysis of Reaction Times in Experiment 2

As was the case in Experiment 1, to investigate whether any differences in recognition accuracy were accompanied by a speed-accuracy trade-off, we examined whether RTs during the recognition task varied with trial type. The mean RTs (and standard errors) for the predictive same-outcome and predictive different-outcome trials were 7.96 log ms (0.05) and 7.96 log ms (0.05) respectively. The mean RT (and standard error) for the non-predictive trials were 8.03 log ms (0.06). A one-way frequentist and Bayesian ANOVAs of individual RTs revealed a main effect of trial type, $F(2, 168) = 3.62$, $p = .029$, $\eta_p^2 = 0.04$, $BF_{inclusion} = 1.02$. The mean RTs on the non-predictive trials were significantly longer relative to the predictive different outcome trials, $t(84) = 2.28$, $p = .025$, $d = 0.25$, $BF_{10} = 1.39$, as well as the predictive same outcome trials, $t(84) = 2.33$, $p = .022$, $d = 0.25$, $BF_{10} = 1.53$. The mean RTs on the predictive same outcome and the predictive different outcome trials did not differ significantly, $t(84) = 0.15$, $p = .885$, $d = 0.02$, $BF_{10} = 0.12$.

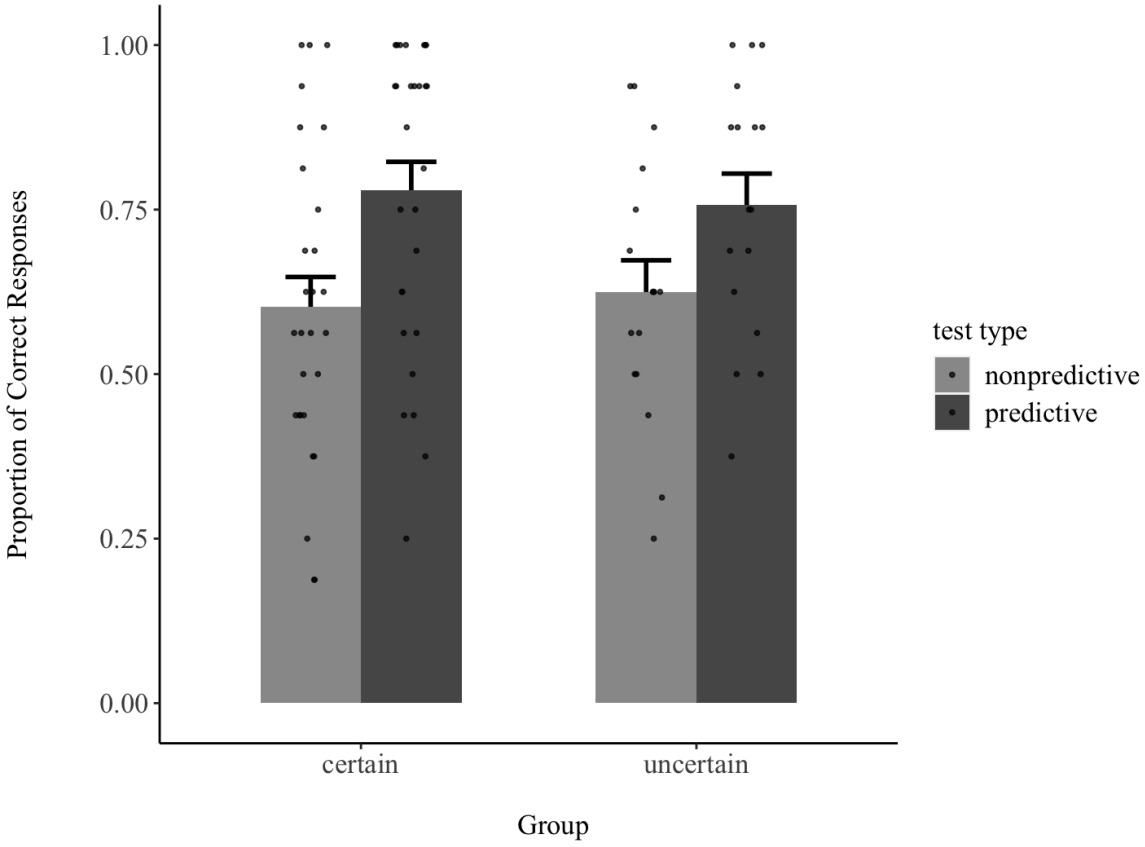
Appendix 4

Analysis of Recognition Accuracy in Experiment 1 adjusted for Training Performance

The original analysis of recognition accuracy revealed that performance was more accurate for the predictive relative to the non-predictive test trials in Group Certain, but not in Group Uncertain. Given that Group Uncertain showed lower accuracy throughout the training phase, we investigated whether the different patterns in recognition accuracy for the two groups might be due to participants who did not learn the distinction between the predictive and the non-predictive cues. Therefore, we re-analysed the recognition accuracy following exclusions of participants with overall training accuracy below .6 (a criterion also used in Le Pelley et al., 2011; Le Pelley & McLaren, 2003). There were 28 (out of 35) participants remaining in Group Certain and 17 participants remaining in Group Uncertain. Figure A1 shows the proportion of correct responses on the predictive and the non-predictive test trials for each participant. A mixed ANOVA revealed a significant main effect of predictiveness: participants made more correct responses on the predictive relative to the non-predictive trials, $F(1, 43) = 16.47$, $p < .001$, $\eta_p^2 = .28$, $BF_{Inclusion} = 414.30$, while the effect of group was not significant, $F(1, 43) < 1$, $p = .995$, $\eta_p^2 = .000007$, $BF_{Inclusion} = 0.32$. In contrast to the original analysis, the interaction between the two factors was not significant, $F(1, 43) < 1$, $p = .566$, $\eta_p^2 = .008$, $BF_{Inclusion} = 0.36$.

Figure A1

Recognition Test Accuracy in Experiment 1 following Exclusions based on the Training Performance: The Proportion of Correct Stimulus Selections during the Predictive and the Non-predictive Test Trials in Groups Certain and Uncertain



Note. Bars represent mean proportion of correct responses; points represent proportion correct responses for individual participants; error bars represent +/-1 SE.

Appendix 5

Reliability Analysis for Experiment 1

The following analysis investigates the reliability of the difference in recognition accuracy between the predictive and the non-predictive trials in Group Certain. The predictive and the non-predictive test trials were numbered separately in order of presentation (1-16); these trial numbers were then labelled as odd or even. For each participant, we calculated the difference between the accuracy score for the predictive and the non-predictive test trials, separately, for the odd and the even trials. There was a significant overall correlation between the odd-trials difference scores and the even-trials difference scores for Group Certain, $r = .42$, $t(33) = 2.66$, $p = .012$ (r = .59 following Spearman-Brown correction). However, as might be expected where participants are showing no consistent difference between conditions, this correlation was not significant for Group Uncertain, $r = .27$, $t(33) = 1.63$, $p = .112$ ($r = .43$ following Spearman-Brown correction).

Reliability Analysis for Experiment 2

The predictive and the non-predictive test trials were numbered separately in the order of presentation (1-16). These trial numbers were then classified as 'odd' or 'even', and the difference in accuracy between the predictive and the non-predictive test trials was calculated separately for the odd and the even trials. There was a significant correlation between these difference scores on the odd and the even trials, $r = .31$, $F(83) = 3.01$, $p = .003$ ($r = .48$ following Spearman-Brown correction).