Durham University

Durham E-Theses

The effects of learned predictiveness and uncertainty on associability

CHANG-MAO CHAO

How to cite:

CHAO, CHANG-MAO (2020) The effects of learned predictiveness and uncertainty on associability. Doctoral thesis, Durham University.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a https://etheses.durham.ac.uk/id/eprint/13482/ is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
https://etheses.durham.ac.uk

The effects of learned predictiveness and uncertainty on associability

Chang-Mao Chao

Thesis submitted for the degree of Doctor of Philosophy

Department of Psychology

Durham University

2019

Declaration

The author confirms that the following works presented in this thesis is his own and proper credits have been given to the works of others in which references have been made. None of the presented works has been submitted elsewhere for any other degree or qualification.

Statement of Copyright

The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

Abstract

The Mackintosh (1975) and Pearce-Hall (1980) models of learning propose that prior experience with a cue affects the attention paid to it, but they differ in their predictions for how attention will change. In human learning tasks, it is commonly found that good predictors of outcomes receive more attention than poor predictors (learned predictiveness). In this thesis, however, the opposite result was found: a cue which leads to an uncertain outcome received more attention (learned uncertainty). Surprisingly, this effect relied on the level of difficulty of training procedure. Attention was allocated to uncertain cues when the training procedure used few uncertain compounds; while attention paid to predictive cues was higher than uncertain cues when the training procedure used more uncertain compounds. These results may reflect that participants engage in different processes of automatic attention and controlled attention based on the difficulty of the task. Moreover, the pattern of these results was also found in the comparison between cues previously used in a biconditional discrimination and predictive cues. When the difficulty of the training procedure was relatively low, biconditional cues received more attention than predictive cues, and the opposite was found when the difficulty of the training procedure was relatively high. Biconditional cues are similar to uncertain cues in that individual cues are uninformative, but they differ from uncertain cues in that they are informative when presented in compound. Therefore, the fact that biconditional cues underwent changes in attention that were similar to uncertain cues suggest that the changes in attention depend on the individual prediction error term rather the than summed prediction error term.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors: Dr. David Sanderson, Dr. Anthony McGregor & Dr. Amanda Ellison. Especially, David, I thank for all his support, patience guidance and tolerance to help me to conduct good experiments and analysis. And his style of supervision is humorous and effective. Also, I thank Dr. Joe Austen for helping my English and the discussion of my studies. Many thanks for my officemates for their support includes participation to my studies and discussion of statistics. Secondly, I would like to express my gratitude to my family and my friends for their support. Thank you for being there and helping me to sort out some practical things. Especially, my wife Yuan, many thanks for all your support and encouragement.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 What is associative learning and attention?

Learning can be defined broadly as a concept that the acquisition of knowledge and skills. Specifically, in this thesis, associative learning was discussed and investigated. Imaging you are driving on the road, when you see the traffic light is red, you probably hit the brakes and stop; you step on the accelerator and start driving when the traffic light is green. Associative learning refers to the hypothetical connection between two or more events (cues and outcomes) that can become linked together. In this example, red light is associated with hitting the brakes and green light is associated with stepping on the accelerator. Perhaps the most famous example is that of Pavlovian conditioning (Pavlov & Anrep, 1928). Conditioning is a process of associative learning in which two kinds of stimuli, the conditioned stimulus (CS) and the unconditioned stimulus (US), are associated together. Based on the observation that dogs salivated (unconditioned response, UR) upon seeing food (unconditioned stimulus, US), Pavlov added a neutral stimulus (the sound of a buzzer, conditioned stimulus, CS) when the food (US) was presented to the dogs. The results showed that dogs began to salivate after hearing the sound of the bell (conditional response, CR). The association between the sound of the bell (CS) and the food (US) had been established.

In the process of associative learning, attention plays an important role (Kruschke, 2003; Mackintosh, 1975; Pearce & Hall, 1980; Pearce & Mackintosh, 2010). Attention is a cognitive process of dealing with information in the daily environment, which helps humans and non-human animals to prioritize specific events or tasks for different conditions. Attention has been shown to influence the process of learning (Broadbent, 1958; Eriksen & James, 1986; Posner, 1980; Treisman, 1964; Treisman & Gelade, 1968; Wolfe, 1994). For example, Broadbent (1958) proposed the attentional filter model. This model suggested that all the information provided by sensory input will enter the sensory buffer. Broadbent suggested that attention acts as a filter that can choose what kind of input should be further processed. Broadbent used a classic dichotic listening task, in which different auditory stimuli were sent to different ears of a person. Participants were instructed to pay attention to the information from one ear (attended channel) and ignore the information from the other ear (unattended channel). Participants were then asked to recall information from both ears. The results showed that participants recalled the information from the attended channel well but failed to recall the information from the unattended channel. A selective filter is designed for choose the information or features from the previous stored because of the limited capacity of human's attention. This study showed that attention is a limited capacity and learning can be driven by attention. Treisman (1964) proposed an attenuation model based on Broadbent's filter model. This attenuation model suggests that we attenuate rather than eliminate the unattended stimuli to process the wanted information. This model added few layers to Broadbent's filter model such as recognition threshold, degree of attenuation and hierarchy of analyzers. Later on, Treisman and Gelade (1980) furtherly developed the feature integration theory, which is one of most significant attention models in human visual attention. The feature integration theory includes two stages: the first stage is the pre-attentive stage, participants analyze the objects to recognize different features (such as colour and shape); the second stage is the focused attention stage, attention plays a role of combining different features into separate perception.

Additionally, the prior experience of learning can shape attentional selection. Specifically, the associability of a cue can be influenced by prior experiences with a cue (Bennett, Wills, Oakeshott, & Mackintosh, 2000; Hall & Pearce, 1979, 1982a; Mackintosh & Turner, 1971). This thesis primarily focuses on the situations in which attention to stimuli is shaped by prior learning experiences. Lashley (1929), and Sutherland and Mackintosh (1971), suggested that whether a stimulus can be learned depends on how much attention is allocated to it. Therefore, how fast the cue-outcome association can be learned is an index of attention. In this way, associability of a cue can be interpreted as how much attention has been allocated to a cue.

Before discussing the relationship between associability and learning, the factors which influence the change of associative strength will be discussed. In conditioning procedures, a CS is repeatedly paired with a US. After these pairings the CS can induce a learned response. Learning theories suggest that the CS-US pairings establish an association between them. An ongoing issue is whether the associative links are determined by the processing of the US or the processing of the CS (associability). US-processing models (e.g., the Rescorla-Wagner model (Rescorla & Wagner, 1972)) explain some effects of learning (e.g., blocking, the prior pairing of CS1-US sabotages the successively CS2 learning of CS1-CS2-US), but fail to explain other learning effects (e.g., latent inhibition, the pre-exposed CS stimulus sabotages the subsequently CS-US association). CS-processing models (Mackintosh, 1975; Pearce & Hall, 1980), however, overcome the weakness to explain those learning effects which US-processing models fail to explain. Rescorla and Wagner (1972) proposed a model, which conceptualizes the associations between the US and the CS. It is a function of the discrepancy between the associative strength of all presented stimuli (ΣV) and the maximum conditioning allowed by the US (λ). The change of associative strength (ΔV) is determined by the difference between the summed associative strength of all presented cues on a given trial (ΣV) and the maximum permitted conditioning of the US (λ), moderated by some learning rate parameters (α and β).

$$\Delta V = \alpha\beta(\lambda-\Sigma V)$$

In this equation, ΔV represents the amount of change in associative strength on a given trial. α is the intensity of the conditioned stimulus. β is the intensity of the unconditioned stimulus. Both α and β are assumed to be fixed values. λ is the maximum level of associative strength determined by the unconditioned stimulus. ΣV is the sum of all the associative strengths of the presented stimuli. ΣV varies based on the training procedures and the current trial. This model suggests that the discrepancy (prediction error) between the predicted US and the actual US determines the extent to which learning occurs. The bigger the prediction error, the greater the change in associative strength. When the discrepancy between the actual US and the expected US is zero, learning will stop. The Rescorla-Wagner model can explain various effects such as blocking (Kamin, 1968) and conditioned inhibition (Rescorla, 1969) by applying the concept that learning is affected by the variation of unconditioned stimulus processing. Explaining the blocking effect is one of the most important contributions of the Rescorla-Wagner model. The blocking effect refers to the finding that the prior pairing of a CS (e.g., CS1) with the US makes the learning of an association between a new CS (e.g., CS2) and the US in subsequent compound training (e.g., CS1-CS2→US) less successful than if the prior CS1→US training had not occurred. Since the CS1→US pairing was pre-exposed to animals, the US in the presentation of CS1-CS2→US was fully predicted by the presence of CS1. Therefore, the value of (λ-ΣV), and therefore ΔV, was close to zero. Consequently, there was no learning. The Rescorla-Wagner model is one of the most influential learning theories that emphasizes learning that depends on US-processing. This model has a simple equation (with few parameters) and can make clear predictions. Moreover, this model contributes significantly to the development of other models in associative learning (Wagner, Brandon, Klein & Mowrer, 1989; Wagner, 2014). However, the Rescorla-Wagner model cannot explain some learning effects, for example, latent inhibition. Latent inhibition occurs when the pre-exposure of a CS retards the CS-US association in the following stage. According to the Rescorla-Wagner model, the associative strength of the pre-exposed CS should be zero, as there is no US presented. Consequently, in the following stage, the CS-US association should not be influenced. But the results showed that the prior learning experiences of pre-exposed CS produces an interference effect. From this example, it could be inferred that associative learning is not determined by the US processing alone. Many other associative theories have suggested that a decrease in the associability of the CS (α, the salience of CS) causes the effect of latent inhibition (Mackintosh, 1975; Wagner, 1981). However, the value of α is assumed to be a fixed value in the Rescorla-Wagner model. Given phenomena such as latent inhibition, the associability of a CS (CS processing) should be considered to be a factor which can influence the strength of associative learning.

The associability of a CS can shape the strength of associative learning. Lawrence (1949, 1950) suggested that animals can learn how to allocate their attention to specific cues. In one of his experiments, rats were trained to learn the association between orientations (turn left or turn right) and dimensions (black-white dimension or rough-smooth dimension). For the black-white dimension the wall of the environment was black or white, and for the rough-smooth dimension the texture of the floor was rough or smooth. In stage 1, rats were trained to learn the discrimination (turn left or turn right) either for a black-white dimension or a rough-smooth dimension. In the subsequent stage, rats were divided into two groups: the relevant group and the irrelevant group. For the relevant group, rats received the same dimension as in the first stage. For the irrelevant group, the different dimension was presented. The results showed that the learning rate of the relevant group was faster than the learning rate of the irrelevant group. In other words, animals learned the discrimination that included previously relevant cues faster than the other discrimination which included previously irrelevant cues. In this case, the prior experiences of the relevance of the learned dimension influenced the associability of a cue.

There are two models providing opposite theories that describe CS processing. The Mackintosh model (1975) proposed that if a given conditioned stimulus (CS) is the best predictor then attention to this CS increases and attention to other stimuli that are not good predictors decreases. According to this model, attention may act as a filter that excludes irrelevant and unnecessary information, which may help the animal focus on the most informative cue. In contrast, in a model proposed by Pearce and Hall (1980), attention to a given CS increases when the CS leads to uncertain outcomes, but attention declines when the CS is perfectly predictive. In this model, Pearce and Hall suggested that animals reallocate their attention to those cues (CSs) which lead to uncertain outcomes (US) so as to reduce the overall uncertainty of the environment. It is worth noting that neither of these proposed models can explain all learning effects. Therefore, several accounts have attempted to integrate these seemingly opposing ideas of CS processing (Le Pelley, 2004; Pearce & Mackintosh, 2010). The next two sections will elucidate the role of attention in associative learning by discussing these two contradictory attentional models.

1.2 The relationship between learning and attention

1.2.1 Learned Predictiveness (Mackintosh model, animal and human supporting studies)

In 1975, Mackintosh proposed a learning theory that described the learning rate of a CS as being shaped by prior learning experiences.

$$\Delta V_A = \alpha_A (\lambda - V_A)$$

$$\Delta\alpha \text{ is positive if } |\lambda - V_A| < |\lambda - V_X|$$

$$\Delta\alpha \text{ is negative if } |\lambda - V_A| >= |\lambda - V_X|$$

V_X is the associative strengths of all other CSs present on a trial. Mackintosh suggested that associative learning relies on how much attention animals pay to the CS. The attention to a CS is α_A in the equation above. The learning rate of a CS is usually considered to be the associability of a CS, or alternatively the attention paid to the CS. When the CS is a good predictor of the US, α_A increases, but when the CS is a poor predictor of the US, α_A decreases. If the given CS is a better predictor of the US than the other CSs, α_A will be close to 1. However, if the given CS is a worse predictor of the US than the other presented CSs, than α_A approaches 0 and little learning will occur. The Mackintosh model describes α based on a comparison of the simultaneously presented conditional stimuli. In other words, relative predictiveness drives the associability by comparing the presented CSs. For instance, if CS1 is a better predictor of the outcome than CS2 then the prediction error of CS1 will be small, while the prediction error of the relatively poor predictor CS2 will be larger. Therefore, CS1 will receive more attention than CS2. The individual prediction errors decide the value of α.

Based on the comparison of individual prediction error, the blocking effect and the overshadowing effect can be explained. To recap, the blocking effect occurs when the prior association between CS1 and the US prevents the formation of an association between CS2 and the US in a following stage in which CS1 and CS2 are presented in compound and predict the US. During the first stage, with the CS1-US pairing, CS1 is a perfect predictor of the US and so the prediction error of CS1 is small and attention paid to CS1 is high. In the following stage, with the CS1 and CS2 compound leading to the US, the prediction error of CS2 is large as the US is perfectly predicted by CS1 already. Therefore, the associability of the CS2 should be low. In a similar way, the Mackintosh model can explain the overshadowing effect. Overshadowing is when two or more CSs (CS1 and CS2) are presented (overshadowing group), and the behavioural control elicited by one CS (CS1) of the compound (CS1-CS2-US) is smaller than the behavioural control gained by CS1 if it had been paired with US alone (control group). The prediction error of CS1 from control group (CS1-US) is low, as CS1 is a perfect predictor of US; while the prediction error of the overshadowing group of CSs (CS1-CS2-US) is relatively high, because CS1 and CS2 are equally predicted by the US. Therefore, attention paid to the CS1 (control group) is greater than attention paid to the overshadowing group of CS1 (CS1-CS2-US).

The Mackintosh model (CS-processing model) can not only explain blocking and overshadowing effects but can also explain latent inhibition, which was beyond the scope of the Rescorla-Wagner model (US-processing model). In the procedure of latent inhibition, the pre-exposure of a CS retards the CS-US association in the following stage. Based on the Rescorla-Wagner model, the associative strength of the pre-exposed CS should be zero. However, according to the Mackintosh model, in the pre-exposed stage (CS → nothing), the pre-exposed CS was no better at predicting nothing than the other CSs present (environment context). Consequently, attention to the CS was low, and remained low when the CS was subsequently paired with the US.

Another learning effect that can be explained by this CS processing model (Mackintosh model) is the intra-extra dimensional shift. Schepp and Schrier (1969) demonstrated the intra-extra dimensional shift in a monkey study. There were two groups: intra-dimensional shifts (ID) and extra-dimensional shifts (ED). For the ID group, monkeys were trained to choose objects by using the same dimension (e.g., shape) across two stages. For the ED group, the presented dimension (e.g., shape) in stage 1 was not relevant to the dimension (e.g., colour) in stage 2. The results showed that animals in the ID group (intra-dimensional shift) had better performance than the animals in the ED group (extra-dimensional shift). That is because that the dimension was perfectly predicted of the outcome for the ID group, thus, the prediction error is small. Then, in the following stage, the same dimension was still perfectly matched with the outcome. The prediction error of the dimension in the second stage was low; While, for the ED group, the dimension of second stage was different form the first stage. Therefore, the prediction error of the dimension was larger in the beginning of the second stage. That's why the learning performance if ID group was better than ED group. This intra-extra dimensional shift has been replicated in other studies (Duffaud, Killcross & George, 2007; George, Duffaud & Killcross, 2010). Duffaud et al., (2007) tested this idea by using the optional-shift procedure. Rats were trained to discriminate a set of stimuli with audio-visual cue compounds in stage 1. As in the study of Schepp and Schrier, one dimension (auditory stimuli or visual stimuli) was relevant and the other was irrelevant. In stage 2, new audio-visual stimuli with both audio and visual dimensions, which were equally linked to solve this discrimination, were presented to all rats. In the test stage, the performance of the relevant dimension group was better than the performance of the irrelevant dimension group. These experiments demonstrated that associative learning can be modulated by the CS processing, in which relevant dimension received more attention than the irrelevant dimension.

So far, the discussed literature relates primarily to non-human work. However, similar associative learning effects have been shown in human studies (Estes, 1969; Estes, 1984; Le Pelley & McLaren, 2003; Le Pelley, Beesley & Griffiths, 2011; Le Pelley, Mitchell, & Johnson, 2013; Livesey, Thorwart, De Fina, & Harris, 2011; Rehder & Hoffman, 2005). For example, Le Pelley and McLaren (2003) used a food-allergy task that tested whether prior learning experience in stage 1 could influence learning during the following stage (see Table 1) in which participants were instructed to predict the allergic reaction (outomes:O1-O4 in Table 1) by the presented foods (cues: different letters in Table 1). In stage 1, cue compounds (e.g., AV in Table 1) were composed of one predictive cue (e.g., cue A) and one non-predictive cue (e.g., cue V) and different compounds led to different outcomes (AV → O1, BV→O2, AW→O1, BW→O2, CX→O2, DX→O1, CY→O2, DY→O1). Cues A, B, C and D were consistently reinforced (predictive cues), with cues A and D always being linked to outcome 1 and cues B and C reliably leading to outcome 2. Conversely, cues V, W, X and Y were partially reinforced (non-predictive cues), because they were linked to outcomes 1 and 2 equally often. In other words, compared to predictive cues (A,B,C,D), cues V,X,W,Y could not provide useful information in terms of choosing one of two outcomes. In stage 2, cue compounds were presented with recombined cues

(AX→O3, BY→O4, CV→O3, DW→O4). All the compounds were certain compounds that either led to outcome 3 or outcome 4, and all the elements were equally linked to either outcome 3 or outcome 4. It is worth to noticing that the novel compounds (EF→3, GH→4, IJ→3, KL→4) were the filler trials, which is not the crucial part in this experiment. Because stage 2 outcomes (outcome 3 and outcome 4) were never presented to participants before, the associative strength of all presented cues to the outcomes in the beginning of stage 2 should be zero. By the end of stage 1, participants experienced the predictive cues (A,B,C,D) are better predictors than non-predictive cues (V,X,W,Y) to stage 1 outcomes. Consequently, in the first trial of stage 2, cues A,B,C,D (previous good predictors) should have strong associative strength than cues V,X,W,Y (previous poor predictors). As a result, through all the stage 2 trials, cues A,B,C,D were the better predictors than cues V,W,X,Y to stage 2 outcomes. For instance, for the trial type AX→O3 (see stage 2 Table 1), A was the better predictor than X to outcome 3. Thus, associability between cue A and outcome 3 was stronger than the associability between X and outcome 3. Based on the Mackintosh model (1975), it could be predicted that cue A and C are better predictors than cue V and X to outcome 3, and cue B and D are better predictors than cue W and Y to outcome 4 by the end of stage 2.

In the test stage, participants rated how likely various cue compounds (AC, BD, VX and WY, see Table 2) were linked to outcome 3 or outcome 4. This test phase provided an index of the associability of the presented cues. The compound AC should be linked to outcome 3 and BD should be linked to outcome 4 based on the stage 2 training. However, if participants paid more attention to predictive cues than to non-predictive cues in stage 1, then in stage 2 the elements A, B, C and D should have received greater attention than the elements V, X, W and Y. In other words, the associability between elements A,B,C,D and stage 2 outcomes was stronger than the associability between elements V,X,W,Y and stage 2 outcomes. Therefore, in the test stage, the predictive compound AC should be linked to outcome 3 and the compound BD should be linked to outcome 4, whereas participants should rate the compounds VX and WY as being equally linked to outcomes 3 and 4. This is exactly what was found in this experiment. Le Pelley and McLaren found predictive cues (A, B, C, D) received more attention than non-predictive cues (V, W, X, Y), providing evidence for the Mackintosh model. This experiment demonstrated that learning can be driven by CS processing in human studies, in which previous predictive cues have stronger associability than non-predictive cues.

Table 1. Experimental design of Le Pelley and McLaren (2003). A-Y = foods; 1-4 = allergic reactions

Stage 1	Stage 2	Test
AV → O1	AX → O3	AC → O3/O4?
BV → O2	BY → O4	BD → O3/O4?
AW → O1	CV → O3	VX → O3/O4?
BW → O2	DW → O4	WY → O3/O4?
CX → O2	EF → O3	EH → O3/O4?
DX → O1	GH → O4	FG → O3/O4?
CY → O2	IJ → O3	IJ → O3/O4?
DY → O1	KL → O4	KL → O3/O4?

Both human and non-human literature shows that prior experiences of a cue (learned predictiveness) can influence the associability of a cue. The associability of a cue is considered as an index of attention in those studies. However, this is different from attention as defined by the traditional cognitive literature (Treisman, 1964; Treisman & Gelade, 1968). Here, it is suggested that attention acts as a filter that helps humans focus on useful information and ignore unwanted information. We might want to know what other factors can be influenced by the learned predictiveness effect (the Mackintosh model) apart from associability. For instance, Le Pelley et al. (2011) used a similar paradigm to Le Pelley and McLaren (2003), with eye-tracking, to investigate the relationship between overt attention and associative learning. In this study, there were two words presented on the screen and participants were instructed to predict which sound would follow each pair of words. Feedback was provided after each trial. Each compound was composed of two words, one of which was predictive and the other was non-predictive in the manner explained for the studies above. Pupillary dwell time on predictive cues was significantly longer than on non-predictive cues, which suggested greater overt attention to predictive cues than to non-predictive cues. These findings are consistent with the Mackintosh attentional model (1975). Combined with previous experiments (e.g., Le Pelley & McLaren, 2003), associability and overt attention are higher for predictive cues than non-predictive cues.

The learned predictiveness effect was discussed in the previous section, in which the amount of attention allocated to predictive cues was greater than for non-predictive cues. A reasonable further question concerns the mechanism of the learned predictiveness effect. Generally, attention can be dividend into two categories: top-down attention and bottom-up attention. Top-down attention refers to an internal control process based on prior experience and the goal of task; while Bottom-up attention refers to the selection process is purely driven by external factors such as the high salience of stimuli (Corbetta & Shulman 2002; Itti & Koch 2001). Both top-down (controlled) attentional control or bottom-up (automatic) attentional control can influence the effect of learned predictiveness. Mitchell, Griffiths, Seetoo & Lovibond (2012) suggested that the effect of learned predictiveness involves top-down attentional control. They used a similar procedure as Le Pelley and McLaren (2003). In stage 1, participants experienced both predictive cues and non-predictive cues. Then, in stage 2, all presented cues were predictive of novel outcomes. There were two groups: the change group and the continuity group. For the change group, participants were instructed that the predictive cues from stage 1 are unlikely to be predictive cues in stage 2. For the continuity group, participants were instructed that the predictive cues from stage 1 continued to be predictive cues in stage 2. The results showed that the learning rate for the change group was very different to the learning rate of the continuity group. During stage 2, for the change group, participants learned previously non-predictive cues better than previously predictive cues. However, for the continuity group, previously predictive cues were well learnt. In other words, by applying different instructions participants consciously reallocated attention to non-predictive cues. Therefore, previously non-predictive cues in the change group received more attention than previously predictive cues. This study demonstrated that the learned predictiveness effect is modulated by top-down attentional control.

Another possibility is that the learned predictiveness effect could be a bottom-up attentional process. Le Pelley, Vadillo and Luque (2013) used a spatial cueing task combined with a categorization task to investigate whether the learned predictiveness effect can be modulated by automatic attention. In their study, the experiment started with a categorization task in which a pair of stimuli including a green square and oblique lines was presented to participants. They were instructed to categorize the paired stimuli into one of two categories (either the green square was predictive or the oblique line was predictive). Feedback was provided on each trial in the categorization task. Half of the participants experienced that the green square was predictive and the oblique lines were non-predictive. The other participants experienced the opposite (the oblique lines were predictive and the green square was non-predictive). Then, in the following stage, a spatial cueing task (dot probe task) was applied to investigate whether the previous categorization task produced the learned predictiveness effect. A pair of stimuli (green square and oblique lines, with either green square on the left and oblique lines on the right or the positions reversed) was presented for 150 ms before disappearing. The location of the pair of stimuli was counterbalanced. A probe (triangle) then appeared in either of the left or right locations. Participants needed to make a manual response by pressing a key corresponded to the location of the probe cue. Different amounts of stimulus-onset asynchrony (SOA) were applied to the spatial cueing task to test whether the learned predictiveness effect was an automatic (bottom-up) process or controlled (top-down) process. SOA was defined as the period between the cueing stimuli and the probe onset. There were two SOA conditions (short SOA of 250 ms and long SOA of 1000 ms) used in the procedure. It was predicted that with the short duration SOA, the prior learning experience (learned predictiveness) would produce a rapid and automatic attentional processing. However, the long duration SOA would allow participants to have enough time to apply top-down attentional control to the task, and this may overcome the automatic attentional processing. Indeed, the results showed that the response time to the location cued by previously predictive stimuli was faster than the location cued by previously non-predictive stimuli only when the SOA was short, but not when it was long. This study suggested that the learned predictiveness effect is modulated by an automatic attentional process (bottom-up) when the period of SOA is short.

1.2.2 Learned Uncertainty (Pearce-Hall model, animal and human supporting studies)

The success of the Mackintosh model is that it provides another explanation of cue competition effects such as blocking and overshadowing. Moreover, it can explain some learning effects (e.g., latent inhibition) which cannot be explained by the Rescorla-Wagner model. However, the Mackintosh model cannot explain the negative transfer effect (Hall & Pearce, 1979; Hall & Pearce, 1982b). In the negative transfer procedure, a CS (tone) was paired with a US (shock) for the experimental group in stage 1, while another CS (light) was paired with the same US (shock) for the control group. Then, in the following stage, a CS (tone) was paired with another US (strong shock) for both groups. According to the Mackintosh model, attention paid to the CS (tone) should be high for the experimental group in stage 1. Therefore, in the following stage, the CS (tone) should receive a high level of attention for the experimental group, but not for the control group, as the tone was not presented to the control group in the first stage. The model predicts that animals in the experimental group should learn faster than the animals in the control group. However, the results of these experiments shows the opposite, animals in the control group learn faster than animals in the experimental group.

Given this situation, Pearce and Hall (1980) proposed another explanation of CS - processing for associative learning:

$$\Delta V = S.\alpha.\lambda$$

where S refers to CS intensity and λ is determined by the intensity of the US. α is the associability of CS which is modified by the following equation:

$$\alpha_n = |\lambda - \Sigma V|_{n-1}$$

α_n (current trial) depends on the absolute value of the difference between the intensity of the US (λ) and the associative strength of all the presented stimuli on the previous trial (n-1). This model makes a different assumption of CS-processing compared to the Mackintosh Model. The Mackintosh model suggests that the individual prediction error ($\lambda - V$) drives associability. The Pearce-Hall model, however, proposed that the summed prediction error determines how much attention is paid to a CS. The Pearce-Hall model suggests that attention to a given CS increases when the CS leads to uncertain outcomes (i.e. when the prediction error ($\lambda - \Sigma V$) is large), but attention decreases when the CS is perfectly predicted (i.e. when the prediction error ($\lambda - \Sigma V$) is small). Animals reallocate their attention to those cues that lead to uncertain outcomes, perhaps to explore an uncertain environment in order to reduce uncertainty.

The Pearce-Hall model has also been supported by many animal studies. As previously discussed, Hall and Pearce found a negative transfer effect which supports their uncertainty principle (Hall & Pearce, 1979). Another example is the restoration of the orienting response (Kaye & Pearce, 1984; Swan & Pearce, 1988; Wilson et al., 1992). In this orienting task, the associative strength of a light was measured under different conditions. There were three groups: partially reinforced group, continuously reinforced group and no reinforcement group. For the partially reinforced group, a CS (light) was presented to animals. The following event was either an unconditional stimulus (food) or no event. For the continuously reinforced group, animals received a US (food) reliably after the CS (light) was presented. For the no reinforcement group, the CS was never paired with food. The results showed that the orienting responses of the continuously reinforced group and the no reinforcement group were smaller than the orienting response of the partially reinforced group. This study showed that leaning can be modulated by the CS processing in which a cue that led to unpredictable events received more attention than the cues which led to predictable events.

In terms of human causal learning, there are a few studies that find in favour of the Pearce-Hall model (Hogarth, Dickinson, Austin, Brown, & Duka, 2008; Griffiths, Johnson & Mitchell, 2011). However, those studies have not been fully replicated. Hogarth et al. (2008) suggested that their data were consistent with the Pearce-Hall model rather than the Mackintosh model. In their procedure, there were two visual cues presented on a monitor and participants were instructed to rate their expectancy of an auditory outcome based on the presented compound. There were three types of compounds. The first type was AX, which was always followed by a noise (AX+, + means noise occurs). The next type was CX, which was never followed by a noise (CX-, - means noise is absent). The third type was BX, which was followed on half of its presentations by a noise (BX+) and on the other presentations by no noise (BX-). In this study, cues A and C were predictive but cue B was non-predictive. According to the Mackintosh model, participants should have paid more attention to cues A and C than to cue B. However, the data showed that the pupil dwell time to cue B was longer than to cues A and C. The results were in line with the Pearce-Hall model in that participants paid more attention to the less predictive cue. Another example is Griffiths et al. (2011), in which a negative-transfer design (Hall & Pearce, 1982b) was combined with a food-allergy task (Larkin, Aitken & Dickinson, 1998) in order to test whether the uncertainty effect could be found in a human causal learning task. The food-allergy task is widely used in human causal learning, in which different foods are considered as cues and are used to predict different allergic reactions (Le Pelley & McLaren, 2003; Shanks & Darby, 1998; Turner, Aitken, Shanks, Sahakian, Robbins, Schwarzbauer & Fletcher, 2004). In the experiment (Griffiths et al., 2011), different cues (A-B) were represented by different foods and allergic reactions represent different outcomes (- means no allergic reaction, + represents minor allergic reaction, ++ indicates critical allergic reaction, see Table 2). There were three groups: novel group, negative transfer group and change group. The experiment consisted of three stages: stage 1, stage 2, and stage 1a which was between stage 1 and stage 2. In stage 1 training, the novel group received B+ and the negative transfer and change groups received A+. In stage 2, all three groups received A++, a critical allergic reaction. Between stage 1 and stage 2, only the change group received A-. The results showed that participants in the novel and change groups learned the new A++ association in stage 2, however the negative transfer group showed slower learning of A++. In stage 1a, the A- trials made cue A uncertain, as A was paired with minor allergic reaction (A+) in stage 1. Therefore, participants from the change group paid more attention to cue A in stage 2. These experiments also demonstrated that learning can be modulated by CS processing, in which the associability of unreliable cues is greater than the associability of reliable cues. Although Hogarth et al. (2008) and Griffiths et al. (2011) found evidence for the Pearce-Hall model, the results have yet to be fully replicated. For instance, Austin and Duka (2010, 2012) tried to replicate Hogarth et al.'s findings by using a similar paradigm. However, Austin and Duka found that cue A (consistently reinforced to the noise outcome) received more attention than cue B (no reinforcement) and cue C (consistently non-reinforced), which was not consistent with the findings of Hogarth et al. The results of the studies of Austin and Duka were not consistent with the Pearce-Hall model. The inconsistencies between the results of these studies warrants further investigation.

Table 2. Experimental design of Griffiths et al. (2011). A and B = foods; - means no allergic reaction, + represents minor allergic reaction, ++ indicates critical allergic reaction. Only the Change Group received the stage 1a.

Group	Stage 1	Stage 1a	Stage 2
Change	A+	A-	A++
Negative Transfer	A+		A++
Novel	B+		A++

Even though there appears to be an absence of robust behavioural evidence in human studies for the Pearce-Hall model, there is still some research revealing the neural correlates of uncertainty in humans (Li, Schiller, Schoenbaum, Phelps & Daw, 2011; Ploghaus, Tracey, Clare, Gati, Rawlins & Matthews, 2000). The definition of prediction error is the discrepancy between the predicted US (V) and the actual US (λ). Thus, prediction error can be formulized as λ-V. If the US is under-predicted, then the value of the prediction error (λ-V) should be positive. However, if the US is over-predicted, then the value of the prediction error should be negative. In this way, the absolute value of the prediction error $|\lambda\text{-}V|$, represents how precisely a cue can predict an outcome, regardless of whether it is under-predicted or over-predicted. In other words, if the relationship between the cue and outcome is well-predicted then the absolute prediction error will be small, but if the relationship between the cue and outcome is uncertain then the absolute prediction error should be large. Human fMRI studies have shown that the human brain encodes the absolute prediction error (Boll, Gamer, Gluth, Finsterbusch & Buchel, 2013; Li et al., 2011; Ploghaus et al., 2000; Vanni-Mercier, Mauguiere, Isnard & Dreher, 2009). For example, Ploghaus et al. utilized an aversive task (differential pain conditioning) and fMRI to examine the predictions of CS learning theories (the Pearce-Hall model). There were three stages in their procedure: the acquisition stage, the counter-expected pain stage and the extinction stage. In the acquisition stage, participants learnt the association between presented signals (different light colours) and thermal stimulation (painful hot, non-painful warm and no stimulation). Then, in the counter-expected pain stage, participants received a painful stimulation during the period of signaling. In the extinction stage, no thermal stimulation was presented after the signaling. The results showed that when the painful stimulation was uncertain (in the extinction stage), the BOLD signal of the left superior parietal gyrus decreased. Since the neural basis of learned uncertainty has been discovered, it is reasonable to assume that the uncertainty effect can be obtained by an appropriate behavioural task. To date, however, there is little robust human behavioural evidence to support the Pearce-Hall model.

1.3 Individual prediction error VS. summed prediction error

So far, two forms of prediction error have been discussed: individual prediction error (the Mackintosh model) and summed prediction error (the Rescorla-Wagner model and the Pearce-Hall model). Both forms of prediction error can explain some learning effects but have their own limitations. For instance, the Rescorla-Wagner model (US-processing) can explain blocking and overshadowing effects but fails to explain latent inhibition as a result of the model only focusing on US-processing. The Mackintosh model (CS-processing) can explain both cue-competition effects and latent inhibition. However, it cannot explain the negative transfer effect. The other CS-processing model, the Pearce-Hall model, can explain the negative transfer effect.

Learning models that incorporate the summed prediction error term suggest that the change in associative strength (ΔV) depends only on the summed associative strengths of the presented cues and not on the individual associative strengths of cues. This hypothesis was tested by Rescorla (2000). In experiment 1, animals were trained to learn the relationship between stimuli and food delivery. In stage 1, there were five trial types: A+ (+ means the outcome occurs), C+, X+, BX- (- means the outcome is absent), DX-. Cues A and C were considered to be excitors as they were perfectly predictive of reinforcement (food). However, cues B and D were inhibitors since they were predictive of the absence of the reinforcer (no food). In the following stage, the trial type AB+ (two cues (A & B) were presented together with an outcome (+)), which was composed of one previous excitor (A) and one previous inhibitor (B), was presented to animals. In the test stage, compounds AD and BC were presented. The compounds presented in the test stage were composed of one excitor (A or C) and one inhibitor (B or D). If the change in associative strength was driven by the summed prediction error then responding to the compounds AC and BD in the test stage should be similar, as the stage 2 training (AB+) should have caused an equal associative change to both A and B. However, if the associative change was determined by the individual error term then stage 2 training (AB+) should have caused different associative changes to A and B. Therefore, the responding to compound AD should not have been equal to the responding to compound BC. The results contradicted the Rescorla-Wagner model, showing that the responding to compound BC was greater than the responding to compound AD. The stage 2 training (AB+) produced a greater associative change of cue B than cue A, suggesting that the associative change is not purely dependent on the summed prediction error.

Similarly, a learning theory that incorporates only an individual prediction error term cannot explain all of the associative learning effects. For instance, the overexpectation effect (Kamin & Gaioni, 1974) cannot be explained by an individual error term. In their procedure of overexpection there were two groups: the experimental group and the control group. In stage 1, A+ and B+ were presented to animals for both the experimental group and the control group. Subsequently, the cue compound AB was paired with the same reinforcement only in the experimental group but not in the control group. During the test stage, the individual cues A and B were presented. If the change of associative strength was driven by the individual error term, the responding to the individual cues (A and B) should be the same for the two groups. In stage 1, the prediction errors of cues A and B are small as they are perfectly predictive of the reinforcement. In stage 2, A and B are still predictive of the outcome and so the individual prediction errors of A and B should remain the same. Therefore, during the test stage, the responding of the experimental group should be similar to the responding of the control group. However, the results showed that the responding to the individual stimuli for the experimental group was attenuated compared to the responding of the control group. As such, learning theories that include only an individual error term are not able to explain all of the learning effects that are observed experimentally. Given the fact that both forms of prediction error cannot perfectly explain all the learning effects, some studies have tried to integrate both individual prediction error and summed prediction error into a new model that might overcome these weaknesses (Le Pelley, 2004; Le Pelley, 2010; Vogel & Wagner, 2017). For example, Vogel and Wagner (2017) combined a summed error term ($\lambda - \Sigma V$) and an individual error term ($\lambda - V$) to formulate another equation, which can explain some empirical data (e.g., Pearce, Dopson, Haselgrove, & Esber, 2012) that the original summed error term model failed to elucidate. However, these integrated models still need further empirical data and more simulations to be convincing replacements for single prediction error models.

The form of prediction error that can determine how much attention participants allocate to a cue is still a subject of debate. Beesley, Nguyen, Pearson & Le Pelley (2015) systematically manipulated the two factors of predictiveness and uncertainty in the same experiment in order to reconcile two contradictory learning theories: the Mackintosh model and the Pearce-Hall model. Participants were instructed to play the role of a scientist to decide which creature (outcome) would be created according to different chemical compounds (cues). There were two conditions: the certain condition and the uncertain condition. In the certain condition, compounds AW, AX, BW and BX were consistently (100%) linked to specific outcomes (AW and AX were followed by outcome 1 and BW and BX were followed by outcome 2). In other words, A and B were predictive cues while W and X were non-predictive cues. In the uncertain condition, compounds CY, CZ, DY and DZ were associated with outcome 1 and outcome 2. CY and CZ were paired with outcome 1 (67%) with a higher probability than outcome 2 (33%), and the reverse was true for compounds DY and DZ. An eye-tracker was utilized to measure pupil dwelling time. The results suggested that participants had greater overt attention for predictive cues than non-predictive cues under the certain condition. This was consistent with the selective attentional processes proposed by Mackintosh (1975). Beesley et al. suggested that participants utilized an attentional exploitation strategy to determine which cue was the best predictor of the outcome. The results also showed that participants spent more time looking at uncertain compounds rather than certain compounds, which was consistent with the Pearce-Hall model (1980). Participants might have been using an attentional exploration strategy to reduce the overall uncertainty of the experiment. This was the first study in associative learning to show both effects of learned predictiveness and learned uncertainty within an experiment, potentially suggesting that the two learning effects might coexist to some extent. However, in Beesley et al.'s study, both effects were measured by overt attention and not by the associability of the cues. Therefore, whether the associability of a cue is driven by summed prediction error (the Pearce-Hall model) or individual error term (the Mackintosh model) remains unsolved.

Apart from those models mentioned above, it should be noted that there are many other models which incorporate with the individual prediction error term or the summed prediction error term. For example, Bush and Mosteller (1951) proposed a US-processing model which involved an individual prediction error term. In this model, the associative change is determined by the discrepancy between the associative strength of the current trial and the asymptote of associative strength of the US. Unlike the Rescorla-Wagner model, this model suggested all of the presented cues were independent of one another. Thus, if there were two cues (X and Y) in the learning stage then the prediction error term for cue X should be λ-V_X and the prediction error term for cue Y should be λ-V_Y. This differs from models that use the summed error term (λ-ΣV).

This thesis will investigate the role of attention in associative learning, especially with respect to the associability of cues. Therefore, the two most influential CS-processing learning theories, the Mackintosh model that incorporates the individual prediction error term and the Pearce-Hall model that includes the summed prediction error term, will be tested in this thesis. Mackintosh's theory suggests that the salience of a CS will increase when it leads to a small prediction error. When a CS leads to a greater prediction error its salience will decrease. Therefore, a good predictor will receive more attention and a bad predictor will receive less attention. The Pearce-Hall model, however, states the opposite. When a cue is partially reinforced then the prediction error will be high and the salience of the cue will be high. Under conditions of continuous reinforcement, the low uncertainty results in a low prediction error and consequently the salience of the cue will be low. Taken together, both the Mackintosh model (predictiveness) and Pearce-Hall model (uncertainty) lead to a high salience of CS, but they differ in the rules for CS processing (α). The Mackintosh model suggests that the individual prediction error ($\lambda - V$) determines how much attention participants pay to a CS, whereas the Pearce-Hall model suggests that the summed prediction error ($\lambda - \Sigma V$) drives the associability.

1.4 Anatomy of the Thesis

In order to test whether the associability of a cue is driven by the summed prediction error term or the individual prediction error term, I conducted a series of experiments that tested two of the most influential learning models: the Pearce-Hall model and the Mackintosh model. Both models lead to a high salience of CS, but the CS processing for each model are different. The Pearce-Hall model suggests that the summed prediction error ($\lambda - \Sigma V$) determines how much attention is paid to a cue, whereas the Mackintosh model proposes that the associability of a cue is modulated by the individual prediction error ($\lambda - V$).

Generally, in human associative learning studies there is support for the Mackintosh model but there is less evidence that supports the Pearce-Hall model. In Chapter 2 (Experiments 1-3), Experiment 1 showed the Mackintosh effect by comparing predictive cues to irrelevant cues. However, in Experiments 2 and 3, the Pearce-Hall uncertainty effect was obtained. Chapter 3 (Experiments 4-8) was designed to investigate the reasons for obtaining the uncertainty effect in Chapter 1, when this result is inconsistent with a proportion of the literature. Chapter 3 also investigated the crucial factor that determines whether learned predictiveness (Mackintosh model) or learned uncertainty (Pearce-Hall model) are observed in a given experiment. The main finding of Chapter 3 was that changing the task difficulty could switch the learning effect between learned predictiveness and learned uncertainty. The learned uncertainty effect was obtained when the training procedures was relatively easy, whereas the learned predictiveness effect was found when the training procedures were relatively difficult. Chapter 3 also studied the nature of what is meant by task difficulty. In Chapter 4 (Experiments 9 - 11), the concept of task difficulty was applied to three experiments to examine the role of attention among different types of cues. The focus for this chapter was on biconditional cue compounds, in which each individual cue is non-informative but the configurations of cues are informative. Compared to predictive cues (for which both the individual error term and the summed error term are low) and uncertain cues (for which the individual error term is low and the summed error term is high), the individual error term for biconditional cues is high but the summed error term is low. Whether the associability of a cue is driven by the summed error term or individual error term can be revealed by comparing these three types of cues.

Chapter 2

Introduction

According to attentional models (e.g., the Mackintosh and Pearce-Hall models), associability is determined by the prediction error term. The Mackintosh model suggests that individual prediction error drives learning effects (learned predictiveness), while the Pearce-Hall model suggests that the summed error term determines how much attention is paid to a cue (learned uncertainty). In this chapter, I will test whether the summed prediction error or the individual prediction error determines how much attention is paid to a cue. Le Pelley & McLaren (2003) used the food allergy prediction task to test the learned predictiveness effect by comparing predictive cues to irrelevant cues (see the general introduction). The results showed that participants rated the AC→O3 and BD→O4 associations more strongly than the VX→O3 and WY→O4 associations. Le Pelley and McLaren suggest that attention paid to predictive cues (A-D) was greater than attention paid to irrelevant cues (V-Y) in stage 1. Therefore, in stage 2, participants paid more attention to previously predictive cues than previous irrelevant cues. These results are consistent with the Mackintosh model: attention paid to predictive cues (with small prediction errors) was higher than irrelevant cues (with large prediction errors). My starting point was to test the reproducibility of this effect using a similar procedure. I used different stimuli (country flags rather than foods) and a different cover story (war scenario rather than allergy prediction). If the learned predictiveness effect is robust, then it should still be found regardless of stimuli and cover story.

In order to observe the uncertainty effect, the coexistence of certain compounds and uncertain compounds might be crucial. Hogarth et al. (2008) found the learned uncertainty effect under a situation in which both certain compounds and uncertain compounds were presented by measuring participants' eye movements. The compound BX was uncertain, while the compounds AX and CX were consistently reinforced. The results showed that the pupillary dwelling time on the uncertain cue B was longer than the dwelling time on the predictive cues A and C. Thus, participants paid more attention to the non-informative cue B. Beesley et al. (2015) also found that uncertainty is a crucial factor in obtaining the learned uncertainty effect. They systematically manipulated the two factors of predictiveness and uncertainty in the same experiment, as the previous section discussed. The results showed that pupil dwell time on the uncertain compounds was greater than certain compounds, which is consistent with the effect of learned uncertainty. The data also suggested that participants spent more time looking at predictive cues than irrelevant cues in the certain compounds, which is evidence for learned predictiveness. Based on these prior works, partially reinforced cue compounds were added to Experiment 2 and Experiment 3 in the current thesis in order to test the reproducibility of the uncertainty effect. In Experiment 2 uncertain cues were compared to predictive cues, while in Experiment 3 uncertain cues were compared to irrelevant cues. Combined with Experiment 1 (in which predictive cues were compared with irrelevant cues), the relative attention received by predictive cues, uncertain cues and irrelevant cues can be compared.

Experiment 2 and Experiment 3 allowed comparisons between two forms of prediction error. In Experiment 2, the comparison was made between uncertain cues, for which the summed prediction error was high and the individual prediction error was high, and predictive cues, for which both the summed prediction error and the individual prediction error were low. In Experiment 3, the comparison was made between uncertain cues and irrelevant cues, for which the summed prediction error was low but the individual prediction error was high. If the cue-outcome association is influenced by the summed prediction error, then uncertain cues (high summed prediction error) should receive more attention than predictive cues (low summed prediction error) and irrelevant cues (low summed prediction error). However, if the associability is determined by the individual prediction error, then predictive cues (low individual prediction error) should receive more attention than uncertain cues (high individual prediction error) and irrelevant cues (high individual prediction error).

Apparatus

All experimental stimuli for all experiments in this chapter were presented on a standard desktop computer with a 19-inch CRT monitor. Presentation of stimuli was controlled by MATLAB with CRS (Cambridge Research System) toolbox and psychtoolbox. The distance between participant and the monitor was 45 cm. Flags of different countries were used as cues. Each flag was 10° x 8° (Width x Height) in size. There were twenty-four country flags: United States, Brazil, Canada, China, United Kingdom, Spain, France, Germany, Israel, Japan, Korea, Mexico, Russia, Singapore, Sweden, Turkey, Benin, Guyana, Jamaica, The Republic of the Congo, Portugal, Cuba, Panama and Uruguay. The outcomes were ‘support’, ‘attack’, ‘retreat’ and ‘surrender’, and were represented by images depicting an apple, bomb, person running and person kneeling, respectively. Each outcome image was 4.6° x 4.3° in size. Participants made responses using a standard computer mouse (see Figure 1, left panel). During the test stage, two flags were presented along with the rating scale (see Figure 1, right panel). The rating scale consisted of the numbers 1-9, and each number was 2° x 2° in size.

Behavioural analysis

There were two parameters recorded across all experiments. Firstly, the accuracy of responses were recorded for stage 1 and stage 2 for each experiment. The accuracy was measured by the proportion of correct responses. Secondly, in the test stage, the ratings were coded such that scores of 1 indicated that the participant rated outcome 3 as very likely and scores of 9 indicated that they rated outcome 4 as very likely. The accuracy analysis is vital, as it provides an index of associability. Due to no feedback in the test phase, the rating scores provided an index showing that the associability between the outcomes and presented compounds. It is worth to mention that reaction time is not critical parameter compared to accuracy. Majority of human learning studies only provided accuracy as an index of associability rather than reaction time (Le Pelley & McLaren, 2003; Le Pelley, et al., 2010; Le Pelley, et al., 2011). Thus, reaction time did not take into account in this thesis.

Terminology

In this chapter, comparisons will be made between predictive, irrelevant and uncertain cues. Predictive cues (e.g. cue A,B,C,D in Table 3, Experiment 1) are those that consistently lead to the same reward. Irrelevant cues (e.g. cue V,X,W,Y in Table 3) are those that themselves are not predictive of the reward but are paired with predictive cues, such that the cue compound (e.g. AV→O1, AW→O1 in Table 3, Experiment 1) is consistently rewarded. Uncertain cues (e.g. cue P,Q,R,S in Table 4, Experiment 2) are those that themselves are not predictive of the reward and are paired with other non-predictive cues, such that the cue compound (e.g PQ→O1/O2, PS→O1/O2, RS→O1/O2, RQ→O1/O2) is not consistently rewarded.

Experiment 1: replication of Le Pelley & McLaren (2003)

Participants:

There were sixteen participants (4 males and 12 females) in the experiments. The age range was 18-31 (mean = 23.7, SD = 4.5). All participants had normal or corrected to normal vision. Durham University Psychology students received participants pool credit and other participants were compensated for their time at a rate of £10/hour.

Procedures:

Participants were instructed to play the role of a soldier in a war scenario and were required to predict which outcome would be correct given the combination of flags presented. There were four outcome pictures: an apple, bomb, person running and person kneeling represented the meaning of support, attack, retreat and surrender respectively. On each trial participants were presented with two country flags that lead to a particular type of outcome. For example, in stage 1, each trial started with the presentation of a compound of two cues (flags) and two outcomes (see Figure 1 upper panel). Based on the combination of flags, participants had to choose either the upper icon (e.g., bomb) or the lower icon (e.g., retreat) by using a left click of the mouse in order to make a response. There was no time pressure for each trial. Immediately after a response was made the feedback “Correct!” or “Incorrect” appeared in the centre of the screen for one second, followed by the next trial starting.

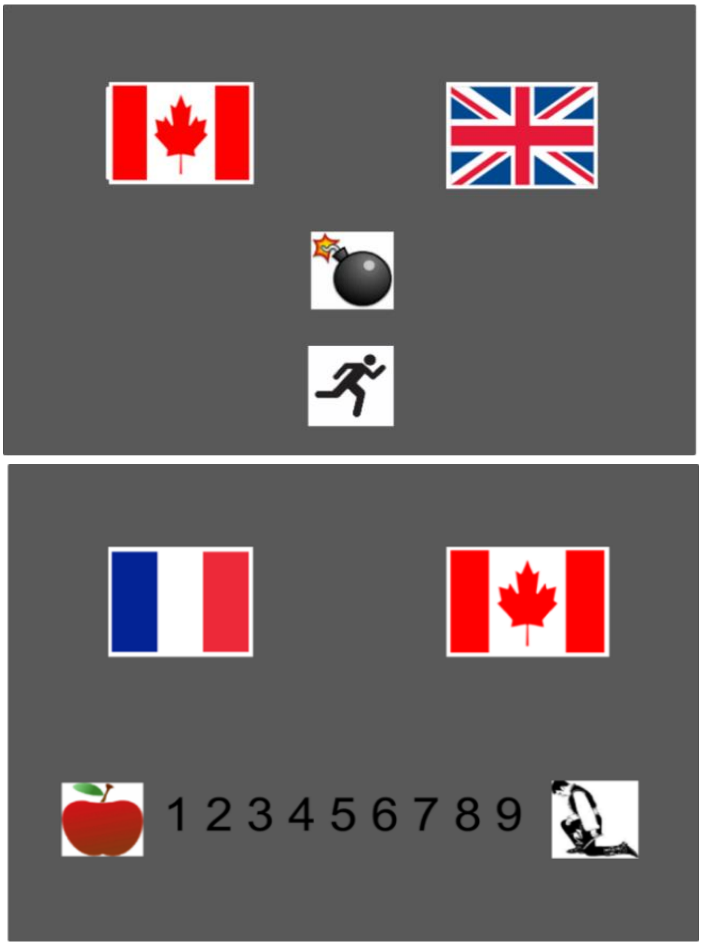

Figure 1. (Upper Panel) Example trial during stage 1. Two cues (flags of Canada and United Kingdom) and two outcomes (pictures representing attack and retreat) were presented. (Lower Panel) Example trial during the test stage. Rating scale (1-9), two cues (flags of France and Canada) and two outcomes (pictures representing support and surrender) were presented.

Table 3. Design of Experiment 1. Letters represent cues (presented in compound) and numbers represent outcomes. During the test stage, participants were asked to rate the expected likelihood of outcome 3 or outcome 4 given the presented cue compound.

Stage 1	Stage 2	Test
AV → O1	AX → O3	AC → O3/O4?
BV → O2	BY → O4	BD → O3/O4?
AW → O1	CV → O3	VX → O3/O4?
BW → O2	DW → O4	WY → O3/O4?
CX → O2	EF → O3	EH → O3/O4?
DX → O1	GH → O4	FG → O3/O4?
CY → O2	IJ → O3	IJ → O3/O4?
DY → O1	KL → O4	KL → O3/O4?

Predictive cues
Irrelevant cues

In stage 1, there were eight trial types (see Table 3) and seven blocks of 16 trials per block. Each block consisted of two trials of each trial type. The order of trials within each block was randomized. In total, there were 112 trials in stage 1. The spatial location of each cue was balanced across the seven blocks, such that each flag was equally presented on either the left or the right of the screen. The location of the correct outcome was randomized. Across trials, one cue in each compound was predictive in that it always led to the same outcome regardless of which cue it was paired with on a given trial (e.g., A is predictive of outcome 1 when presented with other cues: AV→O1, AW→O1). The other cue in each compound was irrelevant by virtue of being paired with two different outcomes equally often (e.g., V is irrelevant when presented with other flags: AV→O1, BV→O2). Therefore, half of the cues (A-D) reliably led to the same outcome and could be considered as good predictors of outcomes. The other stimuli (cues V-Y) were not good predictors as they were partially reinforced (half of the trials on which they were presented led to outcome 1 and half led to outcome 2).

In stage 2, cues were presented in novel compounds and each cue was predictive of one of two novel outcomes: AX→O3, BY→O4, CV→O3, DW→O4, EF→O3, GH→O4, IJ→O3, KL→O4 (see Table 3, stage 2). The first four of these trial types consisted of cue compounds that included one irrelevant cue (V,X,W,Y) and one predictive cue (A,B,C,D) from stage 1 (recombined cues: AX→O3, BY→O4, CV→O3, DW→O4). For the remaining trial types, new cues that had not been previously experienced were used (EF→O3, GH→O4, IJ→O3, KL→O4). These trials with new cues were used as filler trials in order to increase the memory load of stage 2, and replicated, in part, the procedure used by Le Pelley & McLaren (2003). Participants received 64 trials with eight trials of each trial type. The order of trial types was random with the constraint that there were an equal number of each trial type every 16 trials. Cue location was counter balanced and outcome location was randomised in the same manner as stage 1. Because outcome 3 and outcome 4 was never presented to participants in stage 1, there should not be any carryover effect between stages.

In the test stage, novel cue combinations were presented to participants (see Figure 1, lower panel). Given the cue pairings, participants were asked to rate how likely outcome 3 or outcome 4 was on a scale from 1-9. Participants were instructed that choosing either 1 or 9 would indicate that the outcome corresponding to the respective number was very likely, whereas the other outcome was not. For example (see Figure 1, lower panel), if participants selected a rating of 1 on the scale this means that they linked the presented compound to outcome 3. The presented compound was considered to be linked to outcome 4 when participants selected a rating of 9 on the scale. There was no time pressure for each test trial. The next trial appeared immediately after participants rated a score for the given cue compound. There were eight trial types. Half of the trial types (compounds EH, FG, IJ, KL) acted as a control for proper use of the rating scale. As a result of the stage 2 training (EF →O3 and GH→O4), compounds EH and FG should be rated as a score of 5 (not strongly linked to either outcome 3 or 4). Compounds IJ and KL were previously presented in stage 2 and they led to outcomes 3 and 4, respectively. Therefore, participants should rate the compound IJ as strongly linked to outcome 3 and KL as strongly linked to outcome 4. The other half of the compounds (AC, BD, VX, WY) were used to examine the learning effect of stage 2 in order to determine how stage 1 training influenced stage 2 learning. The logic of this procedure followed that of Le Pelley & McLaren (2003). The AC compound consisted of cues (elements A and C) that had both led to outcome 3 in stage 2, and the BD compound consisted of cues (elements B and D) that had led to outcome 4. The VX compound consisted of cues (elements V and X) that had both led to outcome 3 in stage 2, and the WY compound consisted of cues (elements W and Y) that had led to outcome 4.The spatial location of each flag was balanced such that across trials each flag appeared equally often on the left and right. The location of outcome 3 and 4 on the rating scale was random across trials.

Results:

Stage 1: Overall, accuracy in stage 1 increased over blocks and was at approximately 90% by block 7 (see Figure 2). A one-way ANOVA of block (1-7) on accuracy showed a significant effect of block [F(6,90) = 17.85, $p < 0.001$, $\eta_p^2 = .54$, 90% CI [.39, .60], power = 1.00].

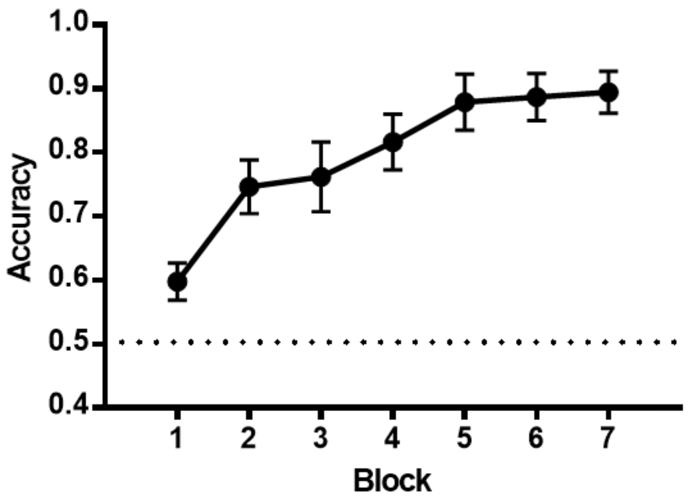

Figure 2. Accuracy across seven blocks of stage 1. Error bars indicate ± the standard error of the mean. The dashed line indicates chance level of 0.5.

Stage 2: Participants acquired the discrimination over training for both the novel cues and the recombined cues from stage 1. Accuracy increased over blocks (see Figure 3). A two-way ANOVA of block (1-4) by condition (recombined compounds AX, BY, CV, DW and control compounds EF, GH, IJ, KL) on accuracy showed a significant main effect of block [$F(3,45) = 86.23$, $p < 0.001$, $\eta_p^2 = .85$, 90% CI [.77, .88], power = 1.00] but no significant main effect of condition [$F(1,15) = 0.21$, $p = 0.65$, $\eta_p^2 = .01$, 90% CI [.00, .20], power = .07]. Moreover, there was no interaction between condition and block [$F(3,45) = 1.1$, $p = 0.36$, $\eta_p^2 = .07$, 90% CI [.00, .16], power = .30].

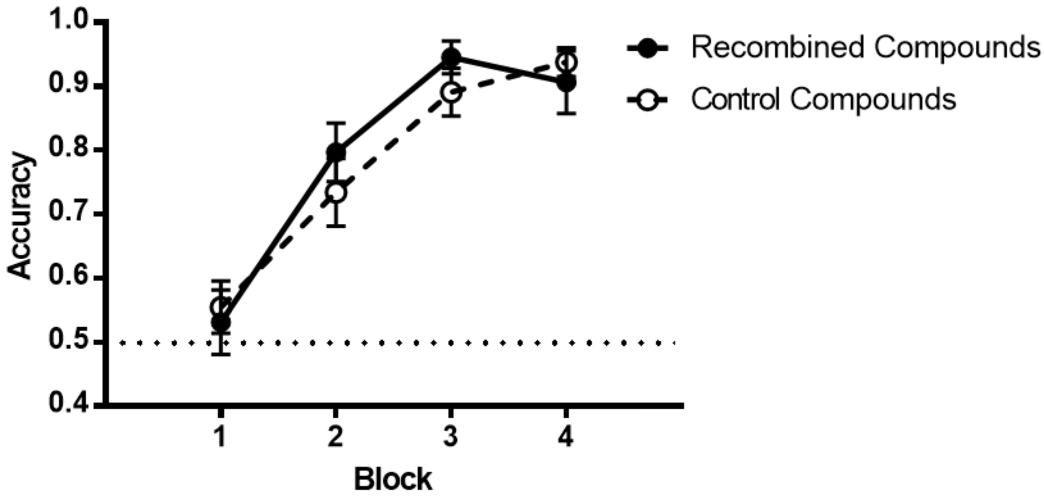

Figure 3. Accuracy across four blocks of stage 2. Error bars indicate ± the standard error of the mean. The dashed line indicates chance level of 0.5.

Test stage: The ratings given for each cue compound during the test stage are shown in Figure 4. The ratings for compounds consisting of cues paired with outcome 4 (BD, WY) were higher than for those paired with outcome 3 (AC, VX), indicating that participants had learnt the cue-outcome associations from stage 2. The difference between cues paired with outcomes 3 and 4 was greater for the predictive compounds than the irrelevant compounds. A two-way ANOVA of condition (predictive cues AC and BD vs. irrelevant cues VX and WY) by outcome (3 for cues AC and VX vs. 4 for cues BD and WY) ANOVA showed a significant main effect of outcome [$F(1,15) = 32.08$, $p < 0.001$, $\eta_p^2 = .68$, 90% CI [.38, .79], power = 1.00], but no significant main effect of condition [$F < 1$, $p = 0.76$]. Importantly, there was a significant interaction between condition and outcome [$F(1,15) = 9.45$, $p = 0.008$, $\eta_p^2 = .39$, 90% CI [.07, .59], power = .86], indicating that the effect of outcome was greater for the predictive cues than the irrelevant cues. Simple main effects analysis showed that there was a significant difference between compounds AC and BD [$F(1,15) = 22.80$, $p < 0.001$, $\eta_p^2 = .60$, 90% CI [.28, .74], power = 1.00], as well as between compounds VX and WY [$F(1,15) = 15.23$, $p = 0.001$, $\eta_p^2 = .50$, 90% CI [.17, .67], power = .97]. Furthermore, the rating for AC was significantly lower than the rating for VX [$F(1,15) = 10.07$, $p = 0.006$, $\eta_p^2 = .40$, 90% CI [.08, .60], power = .88], but the rating for BD was not significantly different from the rating for WY [$F(1,15) = 4.11$, $p = 0.06$, $\eta_p^2 = .22$, 90% CI [.00, .45], power = .52]. One-way ANOVA was carried out to test whether the difference between compound FG and EG, and the difference between compound IJ and KL were significant. There was no significant difference between compound EH and compound FG [$F(1,15) = 0.19$, $p = 0.66$, $\eta_p^2 = .01$, 90% CI [.00, .19], power = .07], but there was a significant difference between the ratings for compounds IJ and KL [$F(1,15) = 31.56$, $p < 0.001$, $\eta_p^2 = .68$, 90% CI

[.38, .79], power = .99]. Two one-sample t-tests showed that the ratings for both compounds (EH and FG) were not significantly different from a rating of 5 [ts < 1, ps > 0.4].

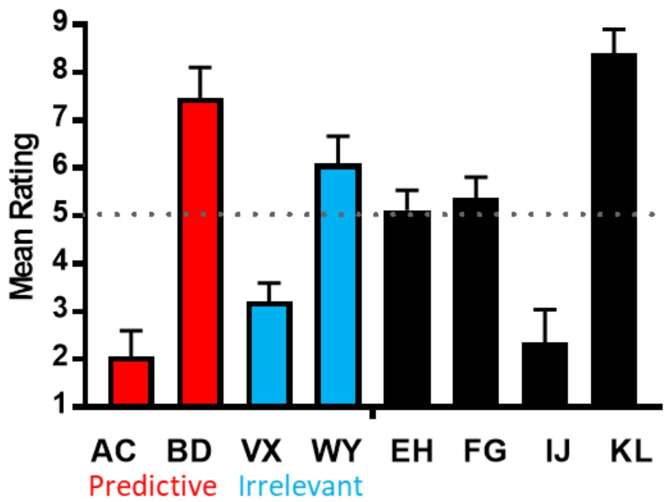

Figure 4. The ratings for each compound in the test stage. The y-axis is the mean rating, with 1 indicating a strong link with outcome 3 and 9 indicating a strong link with outcome 4. The dashed line represents a rating of 5, which indicates that the compound is linked equally with outcomes 3 and 4. Error bars indicate the standard error of the mean.

Discussion:

The main finding of Experiment 1 was that the difference between compounds AC and BD (predictive cues) was larger than the difference between VX and WY (irrelevant cues). This suggests that cues A-D received more attention than cues V-Y in stage 2. This attentional bias is due to the training of stage 1. These results are in agreement with the prediction of the Mackintosh (1975) model that suggests that good predictors (cues A-D) receive more attention than poor predictors (cues V-Y). The individual prediction errors of the predictive cues (A-D) was low, as they were reliable predictors. On the other hand, the individual prediction errors of the irrelevant cues (V-Y) was high. The results of Experiment 1 were a successful replication of Le Pelley and McLaren (2003), using different stimuli and a novel cover story.

It is worth to noticing that I used country flags and a war scenario as visual stimuli and a cover story to replicate the study of Le Pelley and McLaren (2003). It might potentially cause social meaning bias due to the stimuli and cover story. For example, when the cues U.S. flag and U.K flag were presented together and outcomes bomb and apple were presented, participants may possibly choose the apple rather than the bomb. However, due to the randomization of cues and outcomes, any bias effect should be balanced.

Experiment 2: Uncertain vs Predictive – low difficulty

Introduction:

According to the Pearce-Hall model, the summed prediction error ($\lambda-\Sigma V$, the difference between the intensity of the US (outcome) and the associative strength of all the presented stimuli (cues)) drives associability. The certain compounds (AV→1, BV→2, AW→1, BW→2, CX→2, DX→1, CY→2, DY→1) in Experiment 1 had low and similar summed prediction error, as they were consistently paired with specific outcomes. In contrast, the individual prediction error ($\lambda-V$) of predictive cues (A,B,C,D) is lower than irrelevant cues (V,X,W,Y), since predictive cues are the better predictors of outcomes than irrelevant cues. In other words, as a compound (e.g. AV, BV), the summed prediction error is low; while, as an individual cue, the predictive cue (A,B,C,D) has smaller individual prediction error than the irrelevant cue (V,X,W,Y).

For Experiment 2, in order to examine the uncertainty effect, four uncertain compounds were added to the training procedure. The summed prediction error ($\lambda-\Sigma V$) of uncertain compounds is high, while the summed prediction error of certain compounds is low. If the summed prediction error drives associability in this training procedure, as predicted by the Pearce-Hall model, then the attention paid to uncertain cues should be higher than the attention paid to predictive cues. However, the individual prediction error of predictive cues is lower than the individual prediction error of uncertain cues. Therefore, according to the Mackintosh model, the predictive cues should receive more attention than the uncertain cues.

Participants:

Thirty-two participants (6 males and 26 females) completed the experiment. The age range was 18-32 (mean = 23.6, SD = 4.2), and all participants had normal or corrected to normal vision. Durham University Psychology students received participant pool credit, and other participants were compensated for their time at a rate of £10/hour.

Procedures:

The stimuli and cover story were the same as for Experiment 1. In stage 1, participants received eight certain compounds (either predictive or irrelevant) presented in the same way as Experiment 1. They also received four uncertain compounds (PQ, PS, RS, RQ, see Table 4). There were eight certain trial types (either predictive or irrelevant): AV→O1, AW→O1, BV→O2, BW→O2, CX→O2, CY→O2, DX→O1, DY→O2. Cues A-D were predictive and V-Y were irrelevant. For the uncertain compounds, participants were presented with pairs of cues that led to two different outcomes equally often. Participants received four uncertain trial types: PQ→O1/O2, PS→ O1/O2, RQ→O1/O2, RS→O1/O2. There were four blocks in stage 1, and each block contained four trials of each trial type for a total of 48 trials (e.g., four trials for AV → 1, two trials for PQ → 1, two trials for PQ → 2). In total, there were 192 trials in stage 1. All other parameters were the same as Experiment 1.

In the second stage, participants received eight trial types. Four of these were composed of one previously uncertain cue and one previously predictive cue (AP, BQ, CR, DS). The remaining trial types were new cues that were previously not experienced in stage 1 (EF→O3, GH→O4, IJ→O3, KL→O4). These were used as filler trials in order to increase the memory load of stage 2, in a similar manner to Experiment 1.

In the test stage, participants were asked to rate how well the presented compound predicted outcome 3 or outcome 4. The compounds EH, FG, IJ and KL acted as a control for proper use of the rating scale. Compounds AC, BD, PR and QS were the key compounds to test relative attention to uncertain and predictive cues. If attention is equally paid to each element (A, B, C, D, P, Q, R, S) in stage 2, then compounds AC and PR should be rated as outcome 3, and compounds BD and QS should be rated as outcome 4. If, according to the Pearce-Hall model, attention paid to uncertain cues (P, Q, R, S) is higher than predictive cues (A, B, C, D) in stage 1, then, in stage 2, attention paid to the elements P, Q, R and S should remain high. Therefore, the rating difference between compounds PR and QS should be larger than the difference between compounds AC and BD in the test stage. However, if, according to the Mackintosh model, attention paid to predictive cues is higher than uncertain cues, then the difference in ratings for AC and BD should be larger than the difference for PR and QS.

Table 4. Design of Experiment 2. Letters represent cues (presented in compound) and numbers represent outcomes. During the test stage, participants were asked to rate the expected likelihood of outcome 3 or outcome 4 given the presented cue compound.

Stage 1	Stage 2	Test
AV → O1	AP → O3	AC → O3/O4?
BV → O2	BQ → O4	BD → O3/O4?
AW → O1	CR → O3	PR → O3/O4?
BW → O2	DS → O4	QS → O3/O4?
CX → O2	EF → O3	EH → O3/O4?
DX → O1	GH→ O4	FG → O3/O4?
CY → O2	IJ → O3	IJ → O3/O4?
DY → O1	KL → O4	KL → O3/O4?
PQ → O1/O2		
PS → O1/O2		
RS → O1/O2		
RQ → O1/O2		

Predictive cues

Irrelevant cues

Uncertain cues

Results:

Stage 1: Accuracy for certain compound increased across the four blocks of stage 1 but remained at around chance level for uncertain compound (see Figure 5). A two-way ANOVA of block (1-4) by certainty (certain vs. uncertain) on accuracy showed significant main effects of block [$F(3,93) = 10.64$, $p < 0.001$, $\eta_p^2 = .26$, 90% CI [.12, .35], power = 1.00] and certainty [$F(1,31) = 41.48$, $p < 0.001$, $\eta_p^2 = .57$, 90% CI [.36, .69], power = 1.00], and a significant interaction between factors [$F(3,93) = 6.34$, $p = 0.001$, $\eta_p^2 = .17$, 90% CI [.05, .26], power = .97].

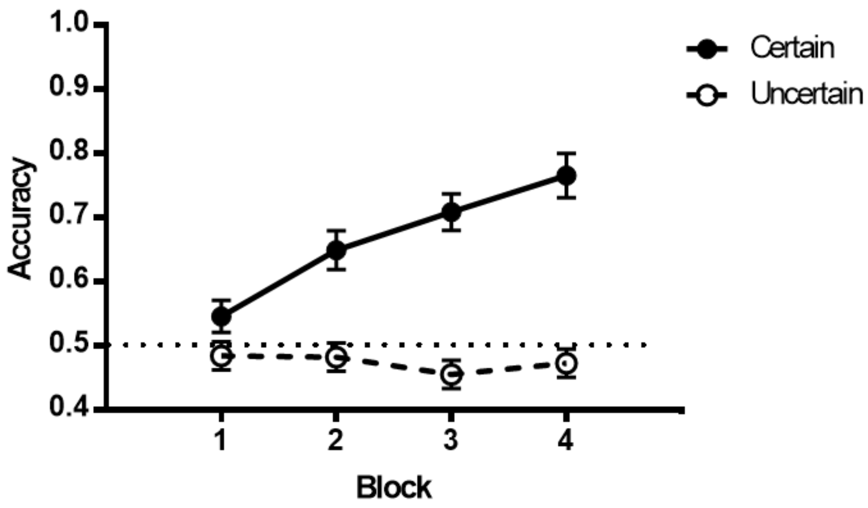

Figure 5. Accuracy across four blocks of stage 1. Error bars indicate ± the standard error of the mean. The dashed line indicates chance level of 0.5.

Stage 2: Participants acquired the discrimination over stage 2 training for both the novel cues (control cues) and the recombined cues from stage 1. Accuracy increased for novel and recombined cues across the four blocks of stage 2 (see Figure 6). A two-way ANOVA of block (1-4) by condition (recombined vs. control) on accuracy showed that there was a significant main effect of block [$F(3,93) = 33.67$, $p < 0.001$, $\eta_p^2 = .52$, 90% CI [.39, .60], power = 1.00], but no significant main effect of condition [$F(1,31) = 2.41$, $p = 0.13$, $\eta_p^2 = .07$, 90% CI [.00, .24], power = .34]. There was no significant interaction between these factors [$F < 1$, $p = 0.41$].

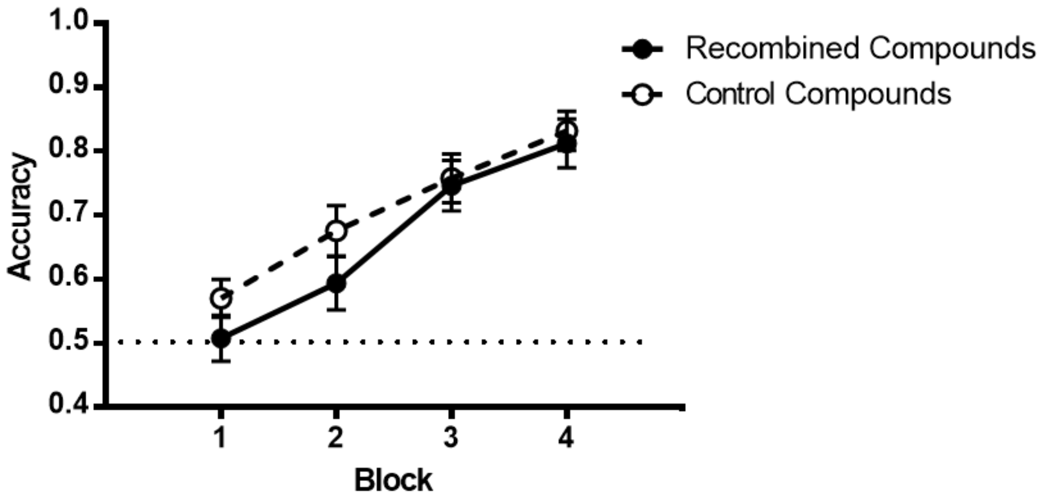

Figure 6. Accuracy across four blocks of stage 2. Error bars indicate ± the standard error of the mean. The dashed line indicates chance level of 0.5.

Test stage: The ratings for cue compounds presented during the test stage are shown in Figure 7. The ratings for compounds consisting of cues paired with outcome 4 (BD and QS) were higher than for those paired with outcome 3 (AC and PR), indicating that participants had learnt the cue-outcome associations from stage 2. More importantly, the difference between compounds PR and QS is greater than the difference between compounds AC and BD. This was confirmed with a two-way ANOVA of condition (predictive cues vs. uncertain cues) by outcome (3 vs. 4), which showed a significant main effect of outcome [$F(1,31) = 21.23$, $p < 0.001$, $\eta_p^2 = .41$, 90% CI [.18, .56], power = 1.00], but no significant main effect of condition [$F < 1$, $p = 0.42$]. Importantly, there was a significant interaction between condition and outcome [$F(1,31) = 11.47$, $p = 0.002$, $\eta_p^2 = .27$, 90% CI [.07, .45], power = .92] demonstrating that the effect of outcome was significantly greater for the uncertain cues than for predictive cues. Simple main effects analysis showed that the score for compound PR was significantly lower than the score for compound AC [$F(1,31) = 10.65$, $p = 0.03$, $\eta_p^2 = .26$, 90% CI [.06, .43], power = .90] and the score for QS was significantly higher than the score for BD [$F(1,31) = 5.55$, $p = 0.025$, $\eta_p^2 = .15$, 90% CI [.01, .33], power = .65]. There was a significant difference in the ratings for the uncertain compounds [$F(1,31) = 31.59$, $p < 0.001$, $\eta_p^2 = .50$, 90% CI [.28, .64], power = 1.00] but a similar comparison for the predictive compounds failed to reach significance [$F(1,31) = 4.16$, $p = 0.05$, $\eta_p^2 = .12$, 90% CI [.00, .30], power = .53]. One-way ANOVA was carried out to test is there any difference between compounds EH and FG and difference between compound IJ and KL. There was no difference between compound EH and FG [$F(1,31) = 2.46$, $p = 0.13$, $\eta_p^2 = .07$, 90% CI [.00, .24], power = .35], but difference between IJ and KL was significant [$F(1,31) = 74.50$, $p < 0.001$, $\eta_p^2 = .71$, 90% CI [.53, .79], power = 1.00], and two one-sample t-tests showed that the ratings for both were not significantly different from a rating of 5 [ts < 1.25, ps > 0.2].

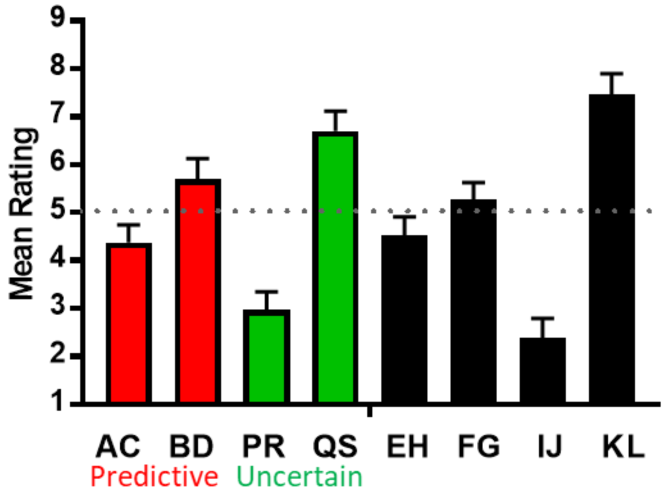

Figure 7. The ratings for each compound in the test stage. The y-axis is the mean rating, with 1 indicating a strong link with outcome 3 and 9 indicating a strong link with outcome 4. The dashed line represents a rating of 5, which indicates that the compound is linked equally with outcomes 3 and 4. Error bars indicate the standard error of the mean.

Discussion:

The difference in mean ratings during the test stage for compounds AC and BD was smaller than the difference between PR and QS. These results suggest that uncertain cues received more attention than predictive cues. As the summed prediction error of uncertain cues was higher than the summed prediction error of predictive cues, these results support the Pearce-Hall model of attentional bias. This uncertainty effect supports the findings of Hogarth et al (2008)., and Griffiths et al (2011). However, this finding is not consistent with some studies (Le Pelley, Turnbull, Reimers & Knipe, 2010; Livesey, et al., 2011). For example, Livesey et al. (2011) used a similar procedure to Le Pelley and McLaren (2003) to examine associability by comparing uncertain cues with predictive cues and uncertain cues with irrelevant cues. Their results showed that predictive cues received more attention than uncertain cues, and uncertain cues received a similar level of attention as irrelevant cues. Therefore, Livesey et al. (2011) is consistent with the Mackintosh model: individual prediction error drives associability. Although the results of Experiment 2 are not in agreement with these previous studies, the procedures were different in several ways, which could help to explain these differences. For example, in the training stage, Livesey et al. had more uncertain compounds than in Experiment 2, which may have increased the memory load and caused different learning effects. Another difference is that the tasks used different stimuli and cover stories. In Livesey et al., simple line drawings of objects were used as cues and pictures of weather were used as outcomes. In contrast, in Experiment 2, country flags were used as cues and different pictures were used as outcomes. The discrepancy between the findings of Experiment 2 and Livesey et al. (2011) will be discussed further in Experiments 4-6.

Experiment 3: Uncertain vs Irrelevant – low difficulty

Introduction:

In Experiment 2, the uncertainty effect was found: uncertain cues received more attention than predictive cues. It may be as a consequence of the summed prediction error of uncertain compounds was higher than the summed prediction error of certain compounds. This finding provides support for the Pearce-Hall model. However, many studies (e.g., Le Pelley, et al., 2010; Le Pelley, Beesley & Griffiths, 2011; Livesey et al., 2011) have found the opposite result: predictive cues receive more attention than uncertain cues. In order to further test the hypothesis that summed prediction error drives associability, Experiment 3 used a similar experimental design in which the summed error term is manipulated in a similar manner of Experiment 2, but with different compounds. Experiment 3 compared the attention gained by uncertain cues and irrelevant cues. Uncertain cues and irrelevant cues are similar in that they are both partially reinforced. In Experiment 3, the difference between them is that irrelevant cues were presented with predictive cues to form compounds with consistent reward (e.g., for the trials AV→O1, BV→O2, AW→O1, BW→O2, A and B are predictive cues whereas V and W are irrelevant cues). Uncertain cues were always presented in compound with other uncertain cues, forming compounds that did not have a consistent reward (e.g., for the trials PQ → O1, PQ→O2, P and Q are uncertain cues). If, according to the Pearce-Hall model, the summed prediction error determines how much attention is paid to a cue, then it could be anticipated that the attention paid to uncertain cues (high summed prediction error) would be larger than the attention paid to irrelevant cues (low summed prediction error). However, if, according to the Mackintosh model, the individual prediction error determines how much attention a cue receives, then the attention paid to uncertain cues (high individual prediction error) and irrelevant cues (high individual prediction error) should be similar.

Participants:

Twenty-four participants (13 males and 9 females) completed the experiment. The age range was 18-38 (mean = 25.4, SD = 4.4). All participants had normal or corrected to normal vision. Durham University Psychology students received participant pool credit, and other participants were compensated for their time at a rate of £10/hour.

Procedures:

Stage 1 was identical to stage 1 of Experiment 2 (see Table 5). There were three types of cue in this stage: predictive cues (A-D), irrelevant cues (V-Y), and uncertain cues (P-Q). Stage 2 training was similar to stage 2 of Experiment 2, but here the recombined compounds each consisted of one uncertain cue and one irrelevant cue (PX→O3, QY→O4, RV→O3, SW→O4, see Table 3). As in the previous two experiments, there were four filler compounds in stage 2 (EF→O3, GH→O4, IJ→O3, KL→O4). The test stage proceeded in a similar manner to the test stage in Experiment 2. Half of the compounds (PR, QS, VX, WY) were used to test whether the learning from stage 1 influenced the attention paid to certain cues during stage 2. The other compounds (EH, FG, IJ, KL) were used as a control for appropriate use of the rating scale. The numbers of trials during each stage of the experiment, and procedural details such as randomisation and counterbalancing were the same as for Experiment 2.

Table 5. Design of Experiment 3. Letters represent cues (presented in compound) and numbers represent outcomes. During the test stage, participants were asked to rate the expected likelihood of outcome 3 or outcome 4 given the presented cue compound.

Stage 1	Stage 2	Test
AV → O1	PX → O3	VX → O3/O4?
BV → O2	QY → O4	WY → O3/O4?
AW → O1	RV → O3	PR → O3/O4?
BW → O2	SW → O4	QS → O3/O4?
CX → O2	EF → O3	EH → O3/O4?
DX → O1	GH→ O4	FG → O3/O4?
CY → O2	IJ → O3	IJ → O3/O4?
DY → O1	KL → O4	KL → O3/O4?
PQ → O1/O2		
PS → O1/O2		
RS → O1/O2		
RQ → O1/O2		

Predictive cues
Irrelevant cues
Uncertain cues

Results:

Stage 1: Participants learned the cue-outcome associations for certain compounds, as the accuracy for these compounds increased over blocks. However, they did not learn the uncertain compounds and accuracy for these compounds remained around chance level (see Figure 8). A repeated measure ANOVA of block (1-4) by certainty (certain vs. uncertain) on accuracy showed significant main effects of block [$F(3,69) = 12.85$, $p < 0.001$, $\eta_p^2 = .36$, 90% CI [.19, .46], power = .99] and certainty [$F(1,23) = 41.48$, $p < 0.001$, $\eta_p^2 = .76$, 90% CI [.58, .84], power = .84] and a significant interaction between these factors [$F(3,93) = 6.34$, $p = 0.001$, $\eta_p^2 = .21$, 90% CI [.06, .32], power = .96].

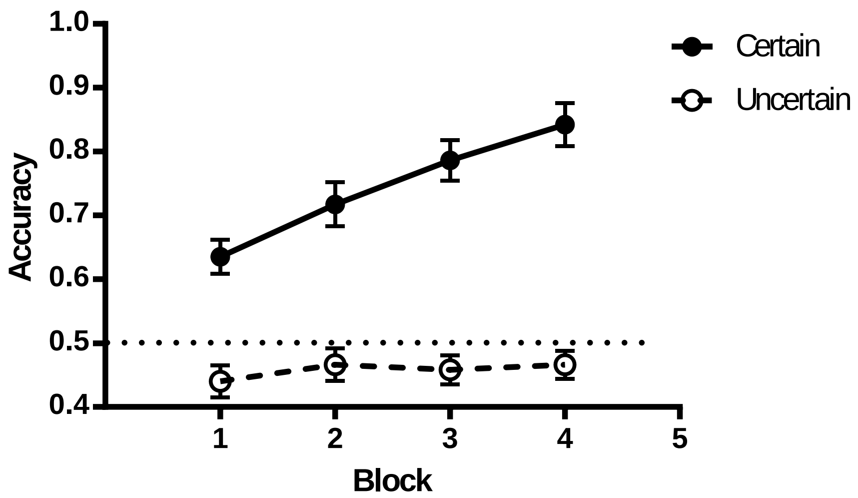

Figure 8. Accuracy across four blocks of stage 1. Error bars indicate ± the standard error of the mean. The dashed line indicates chance level of 0.5.

Stage 2: Participants learned the cue-outcome associations for the control compounds and the recombined compounds, as can be seen from the increase in accuracy over blocks (see Figure 9). An ANOVA of block (1-4) x condition (recombined compounds vs. control compounds) on accuracy showed that there was a significant main effect of block [$F(3,69) = 35.5$, $p < 0.001$, $\eta_p^2 = .61$, 90% CI [.47, .68], power = 1.00], and a significant main effect of condition [$F(1,23) = 18.5$, $p < 0.001$, $\eta_p^2 = .45$, 90% CI [.18, .61], power = .99]. There was no significant interaction between these factors [$F(3,69) = 1.71$, $p = 0.17$, $\eta_p^2 = .07$, 90% CI [.00, .15], power = .45].

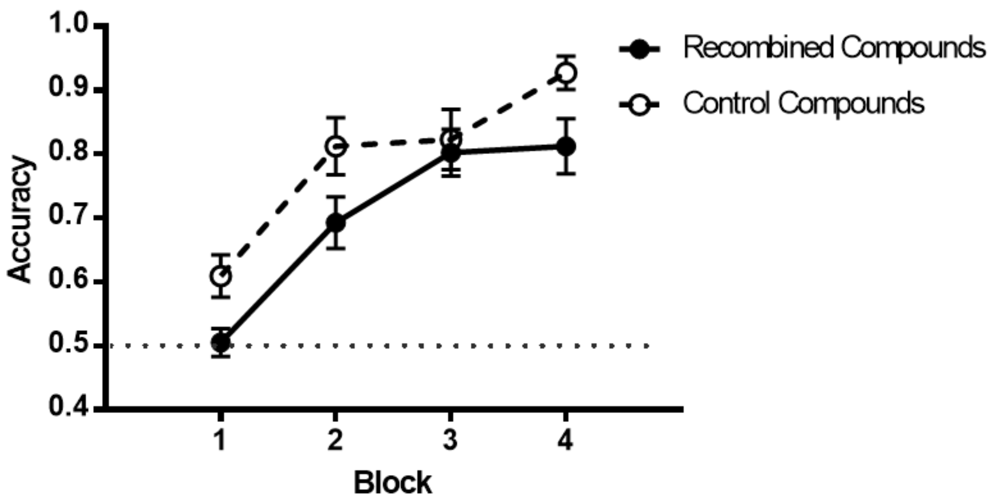

Figure 9. Accuracy across four blocks of stage 2. Error bars indicate ± the standard error of the mean. The dashed line indicates chance level of 0.5.

Test stage: The ratings given to each compound during the test stage are shown in Figure 10. The ratings for compounds consisting of cues paired with outcome 4 (WY and QS) were higher than for those paired with outcome 3 (VX and PR), indicating that participants had learned the cue-outcome associations. Importantly, the difference between ratings for the uncertain cues was larger than the difference between ratings for the irrelevant cues. A two-way ANOVA of condition (irrelevant cues vs. uncertain cues) by outcome (3 vs. 4) showed that there was a significant main effect of outcome [$F(1,23) = 29.11$, $p < 0.001$, $\eta_p^2 = .56$, 90% CI [.30, .69], power = 1.00], but no significant main effect of condition [$F < 1$, $p = 0.43$]. Importantly, the interaction between condition and outcome was significant [$F(1,23) = 18.96$, $p < 0.001$, $\eta_p^2 = .45$, 90% CI [.18, .61], power = .99], confirming that the effect of outcome was greater for the uncertain cues than for irrelevant cues. Simple main effects analysis showed that the score for compound PR was significantly lower than the score for compound VX [$F(1,23) = 4.70$, $p = 0.041$, $\eta_p^2 = .17$, 90% CI [.04 .37], power = .58] and the score for QS was significantly higher than the score for WY [$F(1,23) = 9.6$, $p = 0.005$, $\eta_p^2 = .29$, 90% CI [.06, .49], power = .87]. There were also significant differences between the ratings for the two uncertain compounds PR and QS [$F(1,23) = 33.82$, $p < 0.001$, $\eta_p^2 = .60$, 90% CI [.34, .72], power = .96] and for the irrelevant compounds VX and WY [$F(1,23) = 13.88$, $p = 0.001$, $\eta_p^2 = .37$, 90% CI [.12, .55], power = .96]. One-way ANOVA was conducted to test whether the difference between compound EH and FG, and the difference between compound IJ and KL were significant. There was no significant difference between compounds EH and FG [$F(1,23) = 0.1$, $p = 0.92$, $\eta_p^2 = .01$, 90% CI [.00, .12], power = .06], but there was a significant effect of the difference between compound IJ and KL [$F(1,23) = 86.47$, $p < 0.001$, $\eta_p^2 = .79$, 90% CI [.63, .85], power = 1]. Two one-sample t-tests showed that the ratings for both compounds were not significantly different from a rating of 5 [ts < 1.8, ps > 0.1].

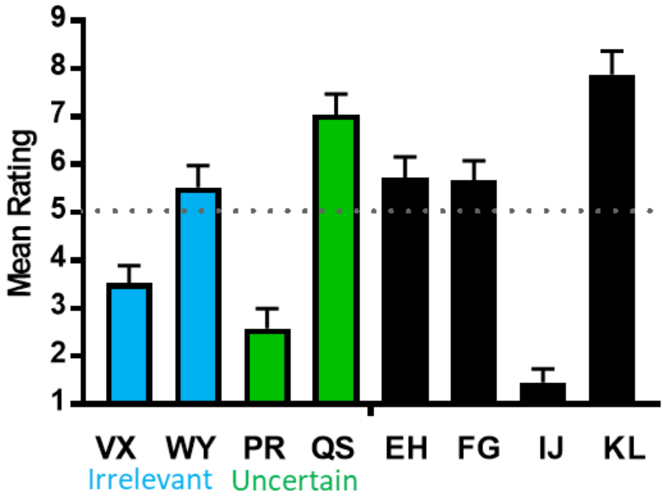

Figure 10. The ratings for each compound in the test stage. The y-axis is the mean rating, with 1 indicating a strong link with outcome 3 and 9 indicating a strong link with outcome 4. The dashed line represents a rating of 5, which indicates that the compound is linked equally with outcomes 3 and 4. Error bars indicate the standard error of the mean.

Discussion:

The main finding of Experiment 3 was that uncertain cues received more attention than irrelevant cues. The summed prediction error of uncertain cues is higher than the summed prediction error of predictive cues; while the individual prediction errors of both types of cues are similar. In this case, the summed prediction error of the uncertain cues was higher than the summed prediction error of the irrelevant cues, and as such the Pearce-Hall model asserts that the uncertain cues should receive more attention. Conversely, the Mackintosh model is unable to account for these results as the individual prediction errors of uncertain cues and irrelevant cues were similar and therefore the attention paid to them should also be similar.

Form Experiment 1 to Experiment 3, the comparisons among predictive cues, irrelevant cues and uncertain cues were made, including predictive cues-irrelevant cues (Experiment 1), uncertain cues-predictive cues (Experiment 2) and uncertain cues-irrelevant cues (Experiment 3). Different forms of prediction errors differed in each type of cue. By running three experiments in this chapter, whether the associability of a cue is driven by the summed prediction error or the individual prediction error can be revealed (see the general discussion).

General Discussion

Experiment 1 replicated Le Pelley and McLaren (2003) with a similar procedure but different stimuli. The results showed that predictive cues received more attention than irrelevant cues, which is in line with attentional processes proposed by Mackintosh (1975). According to this model, predictive cues are good predictors of outcomes, whereas irrelevant cues are poor predictors of outcomes. In this way, the individual prediction error of predictive cues is smaller than the individual prediction error of irrelevant cues and so more attention is paid to irrelevant cues in order to minimize prediction error. Experiments 2 and 3 examined the relative levels of attention gained by different types of cues: predictive cues, uncertain cues, and irrelevant cues. This allowed a direct test of two learning theories with opposing views on attention: the Mackintosh model (governed by individual prediction error) and the Pearce-Hall model (in which summed prediction error drives associability). Experiment 2 made a comparison between predictive cues and uncertain cues; while Experiment 3 compared uncertain cues to irrelevant cues. The results of Experiments 2 and 3 revealed that uncertain cues received more attention than predictive cues and irrelevant cues, providing support for the Pearce-Hall model. The combined results of Experiments 1-3 show a clear hierarchy of attentional bias: uncertain cues received more attention than predictive cues and predictive cues received more attention than irrelevant cues.

The results of Experiments 2 and 3, in which uncertain cues received more attention than both predictive and irrelevant cues, are in agreement with the Pearce-Hall model of summed prediction error. The summed error of uncertain cues is higher than the summed error term of predictive cues and irrelevant cues. In contract, the

Mackintosh model of individual prediction error might not be able to explain these results, as the individual prediction error of predictive cues is lower than uncertain cues and irrelevant cues. However, the results can also be explained by assuming simple learning of the predictive cues. In a situation with the trials AV→O1, AW→O1, BV→O2, BW→O2, a certain compound (e.g., AV) is composed of one predictive cue (e.g., A) and one irrelevant cue (e.g., V). Many studies (e.g., Le Pelley et al., 2011) have shown that predictive cues receive more overt attention than irrelevant cues. As a result, it is possible to solve the task by ignoring the irrelevant cues and focusing solely on the predictive cues. However, this strategy cannot be used for the uncertain compounds, as both cues are uncertain. Therefore, one possible explanation of the results of Experiments 2 and 3 is that participants learned the cue- associations by ignoring the irrelevant cues. This could explain why participants outcome paid more attention to the uncertain cues than the irrelevant cues.

The results of the current experiments (learned uncertainty effects) are not entirely consistent with findings from previous studies (e.g., Le Pelley, et al., 2010; Le Pelley et al., 2011; Livesey et al., 2011). For example, Livesey et al. (2011) found that there was no difference in attention paid to uncertain cues and irrelevant cues. This is consistent with the learning being based on the individual prediction errors of the cues, as the individual prediction errors of uncertain cues and irrelevant cues are high since both cue types are poor predictors. Therefore, their finding is consistent with the Mackintosh model. It is a surprise to find the uncertainty effect, as there is no robust evidence for the learned uncertainty in human learning. There are several differences between my experiments and those studies found the learned predictiveness effects, such as the difference of stimuli and the difference of procedures. These differences and the issue of individual vs. summed prediction error will be further examined in Chapters 3 and 4. Next chapter will attempt to uncover possible reasons why the findings of the current chapter differed from those of previous studies.

Chapter 3:

Introduction

An uncertainty effect was found in Experiments 2 and 3, which is not consistent with previous studies, such as Livesey et al. (2011). This discrepancy could have been caused by my experimental procedures being different from their procedures in various ways. Firstly, the stimuli and cover story of Livesey et al.'s version differed to the stimuli and cover story of my experiments. Secondly, the procedures in Livesey et al.'s study were more complicated than the experimental procedures I used. There were more uncertain compounds in Livesey et al.'s training procedure (four extra uncertain compounds). Thirdly, the cue compounds in the test stage from my experiments is different to the test compound from Livesey et al.'s experiment. The test compounds of Experiment 2 were PR and QS. In stage 1, the uncertain compounds were PR → O1/O2, PS→ O1/O2, RS → O1/O2, RQ → O1/O2. When the test compound PR was presented, it is possible that it reminded participants of cues Q and S, as cue R had been paired with cues Q and S in stage 1. However, in Livesey et al.'s design, this possibility did not exist. Perhaps this distinction caused the difference between my results and those of Livesey et al.

The purpose of this chapter is to investigate the reason why Experiments 2 and 3 showed the uncertainty effect, while Livesey et al.'s experiments showed instead the predictiveness effect. The differences between my experimental procedure and that of Livesey et al. were systematically tested in order to uncover the crucial factor that can drive these differences in learning. In Experiment 4, I used the same procedure as in Experiment 3, but with Livesey et al.'s stimuli and cover story, to examine whether the difference in stimuli caused the observed learning differences. If the uncertainty effect is still observed in Experiment 4, then the difference in stimuli would not be the crucial factor. In Experiments 5 and 6, the number of uncertain compounds was matched with Livesey et al.'s design (increased from four uncertain compounds to eight uncertain compounds) to test whether the number of uncertain compounds determined whether the learning effect observed was predictiveness or uncertainty. In Experiment 5 (see Table 6), the procedure in the test stage was the same as the test procedure in Experiment 2. In Experiment 6 (see Table 7), the test stage procedure was similar to that of Livesey et al. The test compounds were LP and NR. The uncertain training compounds were PQ → O1/O2, PS→ O1/O2, RS → O1/O2, RQ → O1/O2, LM →O1/O2, LO →O1/O2, NO→O1/O2, NM→O1/O2. When the test compounds were presented to participants in Experiment 6, the compound LP did not remind participants of cues N and R, as they had not paired with N and R in the training stage. In the same way, the compound NR did not remind participants of cues L and P. If the observation of learned predictiveness and learned uncertainty was determined by the task difficulty in stage 1, then learned predictiveness will be observed in Experiments 5 and 6. However, if the different learning effects were driven by the test stage procedure, the results of Experiment 5 should be different to the results of Experiment 6. Experiment 5 should show the same effect as Experiment 2, the uncertainty effect, since both experiments have the same test stage procedure; while Experiment 5 should show the learned predictiveness effect as the finding of Livesey et al.

Experiment 4: Replication of Experiment 2 but with Livesey et al.'s (2011) stimuli

Introduction:

In order to test what is the difference between Livesey's experiment and my experiments causing different learning effects (the uncertainty effect was found in chapter 2; the predictiveness effect was found in Livesey's study (2011)), a series of experiments with different manipulations were carried out. In Experiment 4, the key manipulation is to test whether the stimuli and cover story differences can cause the different learning effects. To do that, the same stimuli and cover story as used by Livesey et al. (2011) were applied to Experiment 4, which was otherwise a replication of Experiment 2. Simple line drawings of objects (e.g., bear, airplane) were used as the cues and pictures of weather (rain, snow, hail, fog) were the outcomes. Participants were asked to take part in an invented scenario. They could use magical cards (cues) to control weather (outcomes). This is in contrast to Experiments 1-3, in which country flags were used as the cues and different pictures were utilized as the outcomes. If the differences in stimuli and cover story were the key factor in causing different learning outcomes between my previous experiment and those of Livesey et al., then the effect of learned predictiveness should be observed in Experiment 4. On the contrary, if learned predictiveness and learned uncertainty were not driven by the different stimuli and cover story, then the effect of learned uncertainty should be observed.

Participants:

Thirty-two (20 females, 12 males) people participated in the experiment. The age range was 20-32 (mean: 25.6, SD: 4.6). All participants had normal or corrected to normal vision. Durham University Psychology students received participant pool credit, and other participants were compensated for their time at a rate of £10/hour.

Apparatus

There were twenty object images (cues) and four weather pictures (outcomes). The size of each object image was 10° x 8° (Width x Length), and the size of each outcome was 4.6° x 4.3°. The twenty object images were: airplane, bear, bed, car, clock, doll, elephant, fork, gun, hat, horse, iron, kangaroo, motorbike, pan, pineapple, record player, snowman, telephone, and watering can. The four weather images were: rain, snow, hail, and fog. All experimental stimuli were presented on a standard desktop computer with a 19-inch CRT monitor. MATLAB combined with PsychToolbox and CRS (Cambridge Research System) toolbox were used to control stimuli presentation.

Procedures:

One major difference between Livesey et al.'s (2011) finding and those of Experiments 2 and 3 is the stimuli and cover story. In Experiments 2 and 3, country flags were used as stimuli and the scenario was that participants played the role of a soldier. Livesey et al. used simple line drawing objects as cues and different weather as the outcomes. The cover story they used was that participants were instructed to predict the weather by those simple object pictures. The aim of this experiment was to test whether this difference can cause such opposing results in terms of the learning effects observed. Therefore, the procedure and all other details were the same as for Experiment 2, but the stimuli and cover story were the same as Livesey et al. (2011).

Results:

Stage 1: Participants acquired the discrimination over training with accuracy increasing for the certain compounds over blocks, but no improvement over blocks for the uncertain compounds (see Figure 11). A two-way ANOVA of block (1-4) by certainty (certain and uncertain compounds) on accuracy in stage 1 showed significant main effects of block [$F(3,93) = 6.60$, $p < 0.001$, $\eta_p^2 = .18$, 90% CI [.06, .27], power = .97] and certainty [$F(1,31) = 93.34$, $p < 0.001$, $\eta_p^2 = .75$, 90% CI [.60, .82], power = 1.00], and a significant interaction between block and certainty [$F(3,93) = 6.94$, $p = 0.001$, $\eta_p^2 = .18$, 90% CI [.06, .28], power = .98].

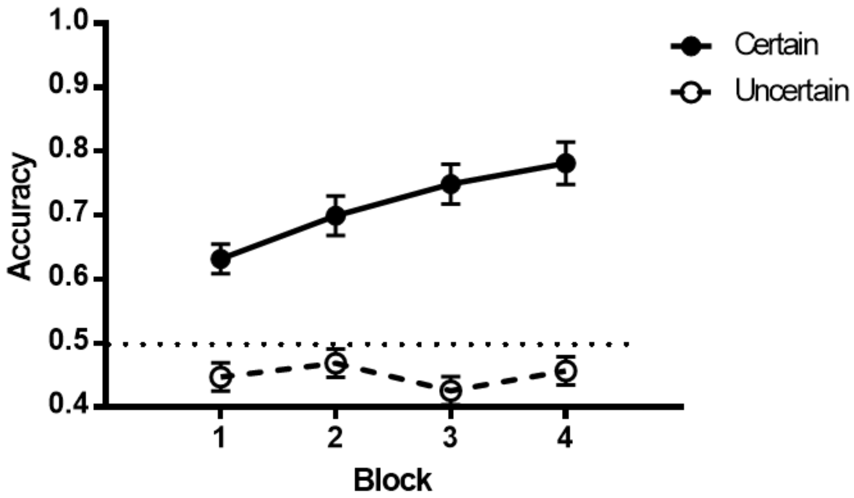

Figure 11. Accuracy across four blocks of stage 1. Error bars indicate ± the standard error of mean. The dash line on the left panel indicates chance level (0.5)

Stage 2: Accuracy increased over blocks for both recombined and control cues (see Figure 12). A two-way ANOVA of block (1-4) by trial-type (recombined compounds and control compounds) on accuracy showed that there was a significant main effect of block [$F(3,93) = 41.77$, $p < 0.001$, $\eta_p^2 = .57$, 90% CI [.45, .64], power = 1.00], but no significant main effect of trial-type [$F(1,31) = 0.17$, $p = 0.69$, $\eta_p^2 = .01$, 90% CI [.00, .11], power = .07]. The interaction between block and trial-type was not significant [$F(3,93) = 2.19$, $p = 0.1$, $\eta_p^2 = .07$, 90% CI [.00, .14], power = .56].

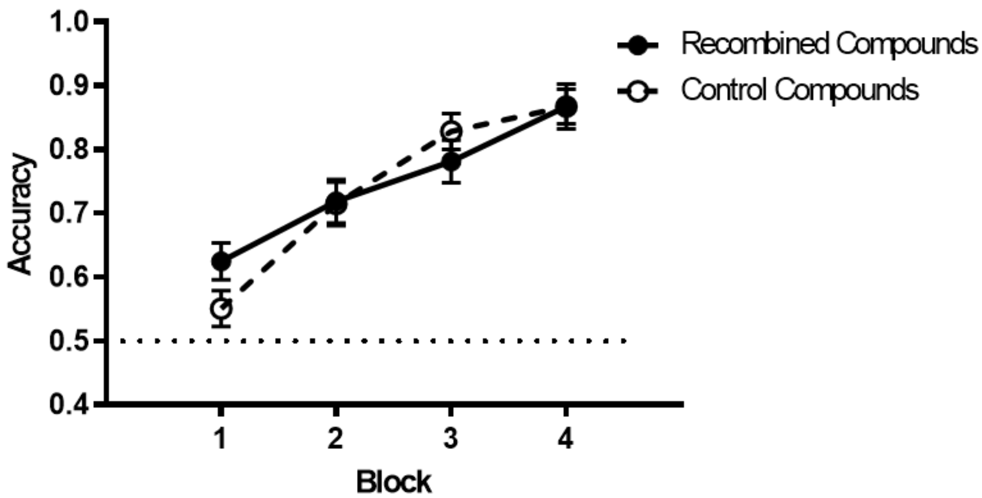

Figure 12. Accuracy in four blocks in stage 2; Error bars indicate ± the standard error of mean. And the dash line on the left panel indicates chance level (0.5)

Test stage: The ratings given for each cue compound during the test stage are shown in Figure 13. Participants were asked to rate how likely the compounds (AC, BD, QR, QS, EH, FG, IJ, and KL) resulted in outcome 3 or outcome 4. The mean ratings of cue compounds paired with outcome 4 (BD and QS) were higher than the mean ratings of cue compounds paired with outcome 3 (AC and PR), suggesting that participants learned the causal relationship of the cue-outcome association. More importantly, the pattern of the rating scales for compounds AC, BD, PR and QS is similar to the pattern of Experiment 2. A two-way ANOVA of condition (predictive cues AC and BD vs. uncertain cues PR and QS) by outcome (compounds AC and PR paired with outcome 3 vs. compounds BD and QS paired with outcome 4) on rating was carried out to test whether previous training experiences (e.g., P,Q,R,S were uncertain cues and A,B,C,D were predictive cues) from stage 1 had an effect on the following stage. The main effect of condition was not significant [$F(1,31) = 0.30$, $p = 0.59$, $\eta_p^2 = .01$, 90% CI [.00, .12], power = .09], but there was a significant main effect of outcome [$F(1,31) = 16.22$, $p < 0.001$, $\eta_p^2 = .34$, 90% CI [.12, .51], power = .98]. Importantly, the interaction between these factors was nearly significant [$F(1,31) = 4.15$, $p = 0.050$, $\eta_p^2 = .12$, 90% CI [.00, .30], power = .53], which suggests that the difference between compounds PR and QS was larger than the difference between compounds AC and BD. Simple main effects analysis showed that the score for compound PR was significantly lower than the score for compound AC [$F(1,31) = 4.92$, $p = 0.034$, $\eta_p^2 = .14$, 90% CI [.01, .32], power = .60] but the score for QS was not significantly higher than the score for BD [$F(1,31) = 1.71$, $p = 0.20$, $\eta_p^2 = .05$, 90% CI [.00, .21], power = .26]. Moreover, there was a significant difference in the ratings for uncertain cues [$F(1,31) = 19.18$, $p < 0.001$, $\eta_p^2 = .38$, 90% CI [.16, .54], power = .99] and for predictive cues [$F(1,31) = 6.06$, $p = 0.02$, $\eta_p^2 = .16$, 90% CI [.01, .34], power = .69].

One-way ANOVA was conducted to test whether the difference between compound EH and FG, and the difference between compound IJ and KL were significant. There was no difference between compound EH and compound FG [$F(1,31) = 0.79$, $p = 0.44$, $\eta_p^2 = .03$, 90% CI [.00, .16], power = .14], and two one-sample t-tests showed that the ratings of both compounds (EH and FG) were not significantly different from a rating of 5 [$ts < 1$, $ps > 0.5$]. However, there was a significant difference between the ratings for compounds IJ and KL [$F(1,31) = 5.30$, $p < 0.001$, $\eta_p^2 = .15$, 90% CI [.01, .33], power = .63].

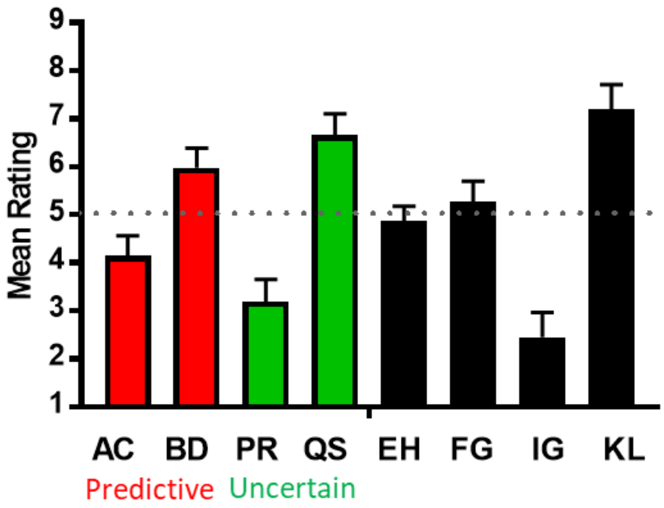

Figure 13. The ratings for each compound in the test stage. The y-axis is the mean rating, with 1 indicating a strong link with outcome 3 and 9 indicating a strong link with outcome 4. The dashed line represents a rating of 5, which indicates that the compound is linked equally with outcomes 3 and 4. Error bars indicate the standard error of the mean.

Discussion:

The main finding of Experiment 4 is that the pattern of the differences between predictive compounds (AC and BD) and uncertain compounds (PR and QS) was similar to that seen in Experiment 2. The results showed that predictive cues received less attention than uncertain cues, an outcome consistent with the uncertainty effect. Once again, this demonstrates a failure to replicate Livesey et al. (2011), in which attention paid to predictive cues was higher than uncertain cues. If the different stimuli and cover story drive different learning effects (learned predictiveness and learned uncertainty), then predictive cues should receive more attention than uncertain cues in Experiment 4. However, combining the results of Experiment 2 (my stimuli) and Experiment 4 (Livesey et al.'s stimuli) together, the uncertainty effect was consistently found. Attention paid to uncertain cues was higher than predictive cues regardless of the nature of the stimuli and cover story. These results indicate that the possibility of different stimuli and cover story causing the contradictory results can be ruled out.

Experiment 5: Uncertain vs. Predictive – high difficulty (the same version as previous experiments (Experiment 2 and 4) except for the number of uncertain compounds)

Introduction:

Experiment 4 showed that uncertain cues received more attention than predictive cues, which is in line with the results of Experiment 2. However, Livesey et al.'s (2011) result showed that attention paid to predictive cues was higher than to uncertain cues. The combined results of Experiments 2 and 4 indicate that the differences in stimuli and cover story cannot be the cause of these opposite learning effects. Another difference between my experiments and Livesey et al.'s study is the number of uncertain compounds present in the training procedure. There were four uncertain compounds in Experiment 2, while there were eight uncertain compounds in Livesey et al.'s procedure. Therefore, for Experiment 5, the number of uncertain compounds was increased to eight, such that this aspect of the design was the same as Livesey et al.'s procedure. Experiment 2 (four uncertain compounds) showed that uncertain cues received more attention than predictive cues. If the number of uncertain cue compounds in the training stage was the crucial factor for this finding, then, in Experiment 5 (eight uncertain compounds), predictive cues should receive more attention than uncertain cues.

Participants:

Twenty-six people (4 males and 22 females) participated in the experiment. The age range was 18-28 (mean: 20.6, SD: 3.1). All participants had normal or corrected to normal vision. Durham University Psychology students received participant pool credit, and other participants were compensated for their time at a rate of £10/hour.

Apparatus

There were twenty-four flags (cues) and four outcome pictures. The size of each flag was 10° x 8° (Width x Length), and the size of each outcome was 4.6° x 4.3°. The twenty-four country flags include: United States, Brazil, Canada, China, United Kingdom, Spain, France, Germany, Israel, Japan, Korea, Mexico, Russia, Singapore, Sweden, Turkey, Benin, Guyana, Jamaica, The Republic of the Congo, Portugal, Cuba, Panama and Uruguay. All experimental stimuli were presented on a standard desktop computer with a 19-inch CRT monitor. MATLAB combined with PsychToolbox and CRS (Cambridge Research System) toolbox were used to control stimuli presentation.

Procedures:

In stage 1, participants received the same cue compounds as in Experiment 2 with an additional four uncertain compounds (NO, NM, ZM and ZO) added to the procedure (see Table 6). These four extra compounds were presented in the same way as the other uncertain compounds (PQ, PS, RS and RQ). This resulted in there being eight certain compounds and eight uncertain compounds in the training stage, which is exactly the same as Livesey et al.'s training procedure. The stage 2 procedure was similar to that of stage 1, with the major difference being the different outcomes (i.e. outcomes 1 and 2 were used in stage 1, but outcomes 3 and 4 were used in stage 2). Each cue was predictive of the outcome (AP→O3, BQ→O4, CR→O3, DS→O4, EF→O3, GH→O4, IJ→O3, KL→O4). Participants received eight trial types in which pairs of cues reliably led to either outcome 3 or 4 (See Table 4, stage 2). Four of these trial types consisted of pairs of cues that included one predictive cue (A,B,C,D) and one uncertain cue (P,Q,R,S) from stage 1 (recombined cues: AP→O3, BQ→O4, CR→O3, DS→O4). For the remaining trial types, new cues that were previously not experienced in stage 1 were used (EF→O3, GH→O4, IJ→O3, KL→O4), which is the same as in previous experiments. In the test stage, participants were asked to rate how likely the presented compounds (AC, BD, PR, QS, EH, FG, IJ, KL) led to outcome 3 or outcome 4. There were eight trial types during the test stage. Half of the trial types (compounds EH,FG,IJ,KL) acted as a control for proper use of the rating scale. The other half of the compounds (AC,BD,PR,QS) were used to examine the learning effect of stage 2 in order to determine whether the stage 1 training influenced stage 2 learning. All other experimental details were the same as for Experiment 2.

Table 6. Design of Experiment 5. Letters represent cues (presented in compound) and numbers represent outcomes. During the test stage, participants were asked to rate the expected likelihood of outcome 3 or outcome 4 given the presented cue compound.

Stage 1	Stage 2	Test
AV → O1	AP → O3	AC → O3/O4?
BV → O2	BQ → O4	BD → O3/O4?
AW → O1	CR → O3	PR → O3/O4?
BW → O2	DS → O4	QS → O3/O4?
CX → O2	EF → O3	EH → O3/O4?
DX → O1	GH→ O4	FG → O3/O4?
CY → O2	IJ → O3	IJ → O3/O4?
DY → O1	KL → O4	KL → O3/O4?
PQ → O1/O2		
PS → O1/O2		
RS → O1/O2		
RQ → O1/O2		
ZM → O1/O2		
ZO → O1/O2		
NO → O1/O2		
NM → O1/O2		

Predictive cues
Irrelevant cues
Uncertain cues

Results:

Stage 1: Accuracy of responding for both certain compounds and uncertain compounds were not significantly increased across blocks (see Figure 14). A two-way ANOVA of block (1-4) by certainty (certain and uncertain) on accuracy of responding showed that there was a significant main effect of certainty [$F(1,25) = 29.68$, $p < 0.001$, $\eta_p^2 = .54$, 90% CI [.29, .67], power = 1.00], but no significant main effect of block [$F(3,75) = 2.46$, $p = 0.07$, $\eta_p^2 = .09$, 90% CI [.00, .17], power = .61] and the interaction between factors was not significant [$F(3,75) = 1.55$, $p = 0.21$, $\eta_p^2 = .06$, 90% CI [.00, .13], power = .41].

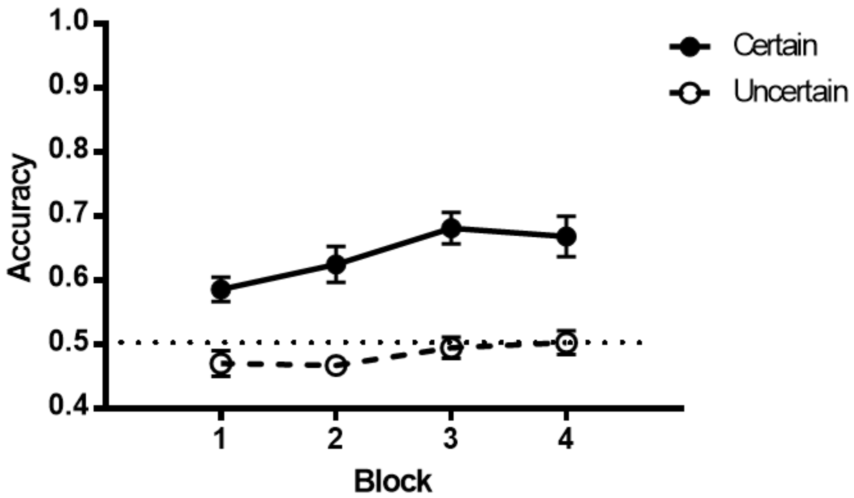

Figure 14. Accuracy across four blocks of stage 1. Error bars indicate ± the standard error of mean. The dash line on the left panel indicates chance level (0.5).

Stage 2: Accuracy increased across blocks for all compounds, but participants were consistently more accurate for control compounds than recombined compounds (see Figure 15). A two-way ANOVA of block (1-4) by trial-type (control compound and recombined compound) on accuracy revealed that there was a significant main effect of block [$F(3,75) = 22.78$, $p < 0.001$, $\eta_p^2 = .48$, 90% CI [.32, .56], power = 1.00], and a significant main effect of trial-type [$F(1,25) = 8.66$, $p = 0.007$, $\eta_p^2 = .26$, 90% CI [.05, .45], power = .84], but there was no significant interaction between these factors [$F < 1$, $p = 0.94$,].

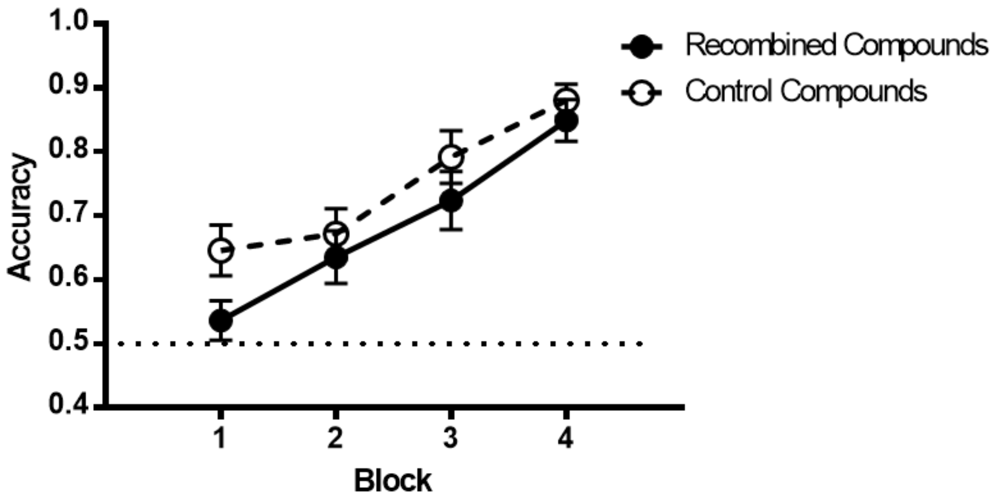

Figure 15. Accuracy in four blocks in stage 2. Error bars indicate ± the standard error of mean. And the dash line on the left panel indicates chance level (0.5)

Test Stage: The ratings for the test stage are shown in Figure 16. The ratings for compounds consisting of cues paired with outcome 4 (BD and QS) were higher than for those paired with outcome 3 (AC and PR), suggesting that participants had learnt the cue-outcome associations. More importantly, the difference between predictive cue compounds (AC and BD) is greater than the difference between uncertain cue compounds (PR and QS), indicating that predictive cues received more attention than uncertain cues. A two-way ANOVA of condition (predictive cues AC and BD vs. uncertain cues PR and QS) by outcome (3 for compounds AC and PR vs. 4 for compounds BD and QS) revealed a significant main effect of outcome [$F(1,25) = 42.12$, $p < 0.001$, $\eta_p^2 = .63$, 90% CI [.40, .74], power = 1.00], but no significant main effect of condition [$F < 1$, $p = 0.76$]. There was a significant interaction between condition and outcome [$F(1,25) = 7.23$, $p = 0.013$, $\eta_p^2 = .22$, 90% CI [.03, .42], power = .77] demonstrating that the effect of outcome was significantly greater for the predictive cues than for the uncertain cues. Simple main effects analysis showed that the rating for compound AC was significantly lower than for compound PR [$F(1,25) = 8.53$, $p = 0.007$, $\eta_p^2 = .26$, 90% CI [.05, .45], power = .84], but the rating for BD was not significantly higher than the rating for QS [$F(1,25) = 3.41$, $p = 0.08$, $\eta_p^2 = .12$, 90% CI [.00, .32], power = .45]. There was a significant difference between the ratings for the predictive cues [$F(1,25) = 36.82$, $p < 0.001$, $\eta_p^2 = .60$, 90% CI [.36, .71], power = 1.00] and for the uncertain cues [$F(1,25) = 8.81$, $p = 0.007$, $\eta_p^2 = .26$, 90% CI [.05, .45], power = .84]. One-way ANOVA was carried out to test is there any difference between compounds EH and FG and difference between compound IJ and KL. There was no significant difference between compounds EH and FG [$F(1,25) = 0.94$, $p = 0.35$, $\eta_p^2 = .04$, 90% CI [.00, .20], power = .16], and two one-sample t-tests showed that the ratings for both compounds were not significantly different from a rating of

5 [ts <1, ps > 0.4]. However, there was a significant difference between compounds IJ and KL [$F(1,25) = 8.9$, $p < 0.001$, $\eta_p^2 = .26$, 90% CI [.05, .45], power = .85].

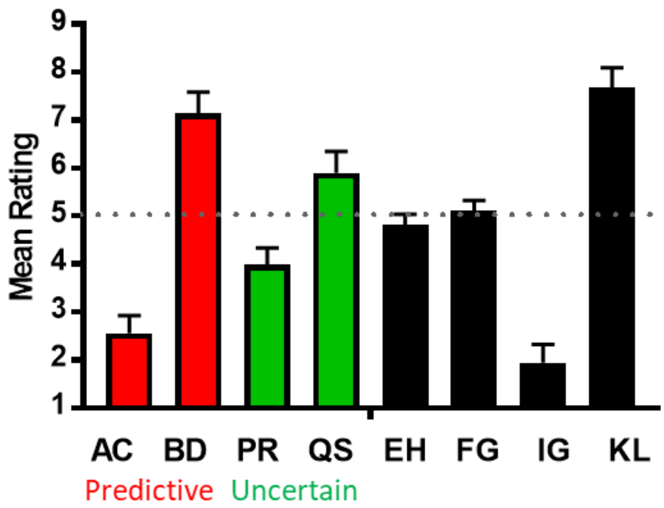

Figure 16. The ratings for each compound in the test stage. The y-axis is the mean rating, with 1 indicating a strong link with outcome 3 and 9 indicating a strong link with outcome 4. The dashed line represents a rating of 5, which indicates that the compound is linked equally with outcomes 3 and 4. Error bars indicate the standard error of the mean.

Discussion:

The important finding of Experiment 5 is that predictive cues received more attention than uncertain cues under the new, relatively more complicated, training procedure. This experiment replicated Livesey et al.'s (2011) finding. Given that the possibility of the contradictory learning effects being caused by differences in the experimental stimuli was ruled out in Experiment 4, there are now only two major differences between Livesey et al.'s experimental design and the design used for Experiment 5. The first difference is the number of uncertain compounds in the training stage, and the other difference is the test stage procedure. Compared to Experiment 2 and Experiment 4 in which the uncertainty effect was observed, Experiment 5 replicated Livesey et al.'s finding of the predictiveness effect by adding four extra uncertain compounds to the training procedure. In other words, when the number of uncertain compounds was four (in Experiments 2 and 4), the results showed that uncertain cues received more attention than predictive cues (learned uncertainty), but when the number of uncertain compounds was eight (Experiment 5), the results showed that attention paid to predictive cues was higher than attention paid to uncertain cues (learned predictiveness).

These results are hugely surprising. By manipulating the difficulty of the training procedure, either the learned predictiveness or the learned uncertainty effect can be found. Experiments 2 and 4 provided novel findings in which uncertain cues received more attention than the other cues. However, the results of Experiment 5 were more in line with previous findings (e.g., Livesey et al., 2011), indicating that the difficulty of the training procedure is crucial. Therefore, this series of experiments seems to provide a solution for an important theoretical question. However, there are still some potential issues that need to be clarified. Firstly, the test stage procedure in Experiment 5 differed from the test procedure in Livesey et al.'s design. It is important to know if the predictiveness effect seen in Experiment 5 remains if the test procedures are matched. Experiment 6 will address this question. Secondly, the uncertainty effect (Experiment 2 and Experiment 4) and the predictiveness effect (Experiment 5) were observed in separate experiments. Therefore, two paradigms (Experiments 2 and 5) were combined to form Experiment 7, in which the learning effects are directly compared. Thirdly, the complexity of the training procedure is an unclear concept. In Experiments 2 and 4 there were eight certain compounds and four uncertain compounds, while in Experiment 5 there were eight certain compounds and eight uncertain compounds. The extra four uncertain compounds in Experiment 5 not only increased the number of uncertain compounds but also increased the memory load. It could be either of these factors that affected the observed learning effect. Experiment 8 will further discuss the exact nature of the complexity of the training procedure.

Experiment 6: Uncertain vs. Predictive – high difficulty (Livesey et al.'s (2011) version)

Introduction:

Experiment 4 ruled out the possibility that the differences in stimuli and cover story between my experiments and those of Livesey et al. (2011) were the cause of the opposing results, as both Experiments 2 and 4 showed that uncertain cues received more attention than predictive cues. In Experiment 5, in which the training procedure was made more difficult, the outcome (learned predictiveness) was the opposite to that seen in Experiment 2 (learned uncertainty). Therefore, it would seem like the complexity of the training procedure was the key factor in determining the learning effect observed. However, the test stage procedure was not identical between Experiment 5 and Livesey et al.'s procedure. In Experiment 5, the uncertain test compounds were PR and QS. The test compound PR could possibly remind participants of cues Q and S, as cue R had been paired with cues Q and S in stage 1 (PR → O1/O2, PS→ O1/O2, RS → O1/O2, RQ → O1/O2). However, in Livesey et al.'s design, this possibility did not exist. It is possible that this difference was the cause of the differences in outcomes between these experiments. Therefore, the test procedure of uncertain compounds from Livesey et al.'s study was applied to Experiment 6. In stage 1 of Experiment 6, the uncertain training compounds are PQ → O1/O2, PS→ O1/O2, RS→ O1/O2, RQ→ O1/O2, ZM→ O1/O2 ZO→ O1/O2, NO → O1/O2, NM→ O1/O2, and in the test stage the uncertain test compounds are PZ and NR. If Experiment 6, with eight uncertain compounds during training and Livesey et al's uncertain compounds during the test, still results in predictive cues receiving more attention than uncertain cues, then it could be concluded that the complexity of the training procedure determines which types of cue receive most attention, rather than the uncertain compounds used during the test stage.

Participants:

Twenty-four people (9 males and 13 females) participated in the experiment. The age range was 18-31 (mean: 22.0, SD: 3.7). All participants had normal or corrected to normal vision. Durham University Psychology students received participant pool credit, and other participants were compensated for their time at a rate of £10/hour.

Apparatus

All the details are the same as for Experiment 5.

Procedures:

The experimental design of stage 1 was exactly the same as Experiment 5, which includes eight uncertain compounds (PQ → O1/O2, PS→ O1/O2, RS→ O1/O2, RQ→ O1/O2, ZM→ O1/O2, ZO→ O1/O2, NO → O1/O2, NM→ O1/O2). In stage 2, four of the trial types (AP→O3, BQ→O4, CR→O3, DS→O4) consisted of pairs of cues that included one predictive cue (A,B,C,D) and one uncertain cue (N,P,R,Z) from stage 1. For the remaining trial types (EF→O3, GH→O4, IJ→O3, KL→O4), new cues that were previously not experienced in stage 1 were used, which is the same as in previous experiments. In the test stage (see Table 7), compounds (AC,BD,PZ,NR,EH,FG,IJ,KL) were rated by participants as to which outcome they should be most associated with. Compared to the test compounds of Experiment 5 (AC,BD,PR,QS,EH,FG,IJ,KL), the major difference of the test compounds between Experiment 5 and Experiment 6 is the two test compounds (PZ and NR). All other details were the same as the previous experiments.

Table 7. Design of Experiment 6. Letters represent cues (presented in compound) and numbers represent outcomes. During the test stage, participants were asked to rate the expected likelihood of outcome 3 or outcome 4 given the presented cue compound.

Stage 1	Stage 2	Test
AV → O1	AP → O3	AC → O3/O4?
BV → O2	BR → O4	BD → O3/O4?
AW → O1	CZ → O3	PZ → O3/O4?
BW → O2	DN → O4	NR → O3/O4?
CX → O2	EF → O3	EH → O3/O4?
DX → O1	GH → O4	FG → O3/O4?
CY → O2	IJ → O3	IJ → O3/O4?
DY → O1	KL → O4	KL → O3/O4?
PQ → O1/O2		
PS → O1/O2		
RS → O1/O2		
RQ → O1/O2		
ZM → O1/O2		
ZO → O1/O2		
NO → O1/O2		
NM → O1/O2		

Predictive cues
Irrelevant cues
Uncertain cues

Results:

Stage 1: Accuracy of responding for both certain compounds and uncertain compounds were not significantly increased across blocks (see Figure 17). A two-way ANOVA of block (1-4) by certainty (certain and uncertain compound) on accuracy showed that there was a significant main effect of block [$F(3,69) = 8.52$, $p < 0.001$, $\eta_p^2 = .27$, 90% CI [.11, .38], power = .99] and a significant main effect of certainty [$F(1,23) = 30.37$, $p < 0.001$, $\eta_p^2 = .57$, 90% CI [.31, .70], power = 1.00], but there was no significant interaction between these two factors [$F(3,69) = 1.48$, $p = 0.23$, $\eta_p^2 = .06$, 90% CI [.00, .14], power = .40].

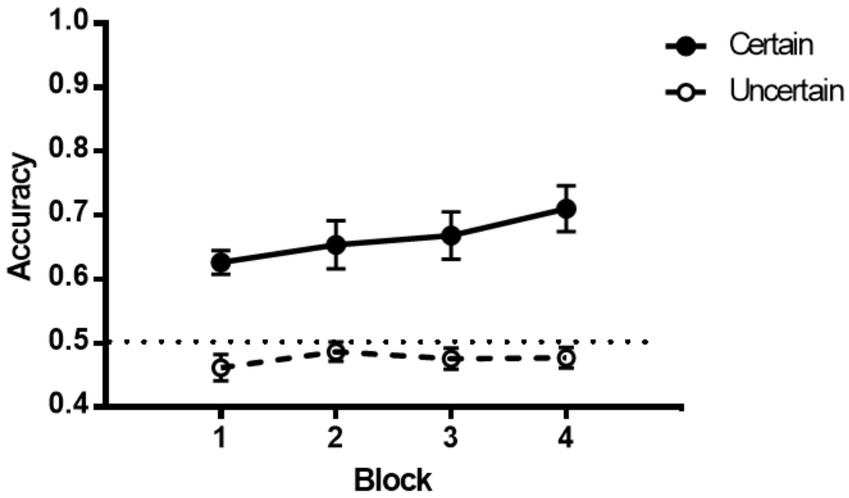

Figure 17. Accuracy across four blocks of stage 1. Error bars indicate ± the standard error of mean. The dash line on the left panel indicates chance level (0.5)

Stage 2: Accuracy increased over blocks for all compounds and was higher for control compounds than for recombined compounds (see Figure 18). A two-way ANOVA of block (1-4) by trial-type (control compound and recombined compound) on accuracy showed that there was a significant main effect of block [$F(3,69) = 26.64$, $p < 0.001$, $\eta_p^2 = .54$, 90% CI [.38, .62], power = 1.00], and a significant main effect of trial-type [$F(1,23) = 4.54$, $p = 0.045$, $\eta_p^2 = .16$, 90% CI [.00, .37], power = 75]. There was no significant interaction between factors [$F(3,69) = 1.21$, $p = 0.31$, $\eta_p^2 = .05$, 90% CI [.00, .12], power = .33].

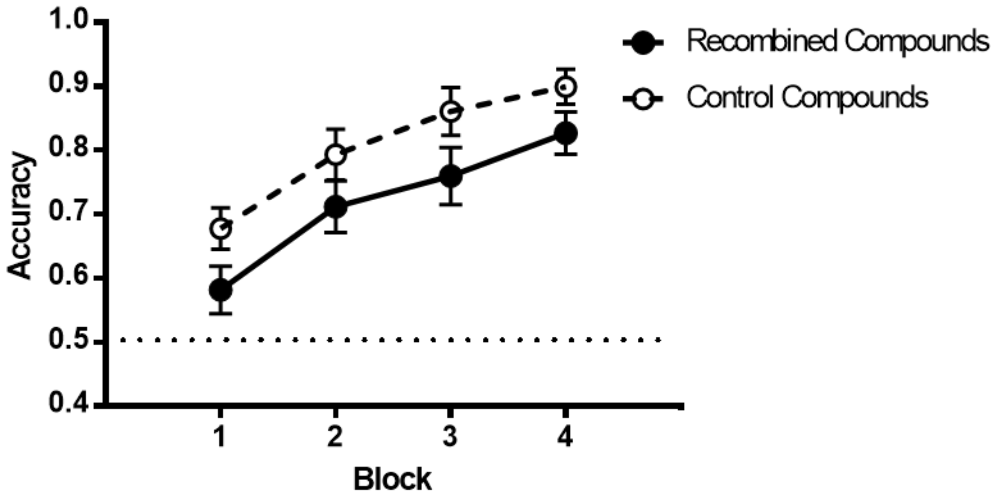

Figure 18. Accuracy in four blocks in stage 2. Error bars indicate ± the standard error of mean. And the dash line on the left panel indicates chance level (0.5)

Test stage: The ratings for each compound during the test stage are shown in Figure 19. There was a larger difference between ratings of the predictive cues (AC and BD) than of the uncertain cues (PZ and NR). A two-way ANOVA of condition (predictive cues AC and BD vs. uncertain cues PZ and NR) by outcome (outcome 3 for compounds AC and PZ vs. outcome 4 for compounds BD and NR) on ratings showed that there was a significant main effect of outcome [$F(1,23) = 41.53$, $p < 0.001$, $\eta_p^2 = .64$, 90% CI [.41, .75], power = 1.00], but there was no significant main effect on condition [$F(1,23) = 3.62$, $p = 0.07$, $\eta_p^2 = .14$, 90% CI [.00, .34], power = .48]. The interaction between these factors was significant [$F(1,23) = 13.85$, $p = 0.001$, $\eta_p^2 = .38$, 90% CI [.12, .55], power = .96], demonstrating that the difference between cue compounds AC and BD was larger than the difference between cue compounds PZ and NR. These results show that the amount of attention paid to predictive cues (A,B,C,D) was higher than the amount of attention paid to uncertain cues (P,N,R,Z). Simple main effects analysis showed that the rating for compound PZ was not significantly different from the rating for compound AC [$F(1,23) = 0.58$, $p = 0.45$, $\eta_p^2 = .03$, 90% CI [.00, .19], power = .12], but the rating of NR was significantly lower than the rating of BD [$F(1,23) = 19.72$, $p < 0.005$, $\eta_p^2 = .46$, 90% CI [.19, .62], power = .99]. There was a significant difference between ratings for the uncertain compounds [$F(1,23) = 12.04$, $p = 0.002$, $\eta_p^2 = .34$, 90% CI [.09, .52], power = .93] and for the predictive compounds [$F(1,23) = 46.47$, $p < 0.001$, $\eta_p^2 = .67$, 90% CI [.44, .77], power = 1.00]. One-way ANOVA was carried out to test is there any difference between compounds EH and FG and difference between compound IJ and KL. There was no significant difference between compounds EH and FG [$F(1,23) = 0.66$, $p = 0.52$, $\eta_p^2 = .03$, 90% CI [.00, .19], power = .13], and two one-sample t-tests showed that the ratings of both compounds (EH and FG) were not significantly different from a rating of 5 [ts < 1, ps > 0.45]. However, there was a significant difference between compounds IJ and KL [$F(1,23) = 5.67$, $p < 0.001$, $\eta_p^2 = .20$, 90% CI [.01, .40], power = .66].

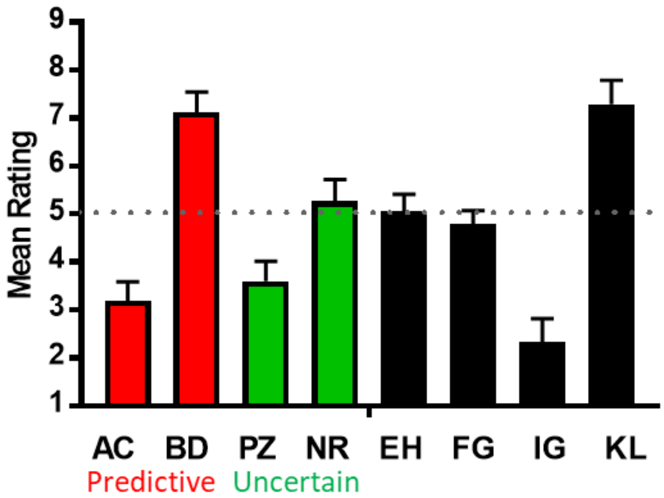

Figure 19. The ratings for each compound in the test stage. The y-axis is the mean rating, with 1 indicating a strong link with outcome 3 and 9 indicating a strong link with outcome 4. The dashed line represents a rating of 5, which indicates that the compound is linked equally with outcomes 3 and 4. Error bars indicate the standard error of the mean.

Discussion:

The main finding of Experiment 6 is that the difference between compounds AC and BD was larger than the difference between compounds PZ and NR, suggesting that uncertain cues received less attention than predictive cues. Experiments 5 and 6 shared the same training procedure in stage 1 (PQ → O1/O2, PS→ O1/O2, RS→ O1/O2, RQ→ O1/O2, ZM→ O1/O2, ZO→ O1/O2, NO → O1/O2, NM→ O1/O2), but the procedures were different in the test stage. There were two forms of uncertain compounds in the test stage: the compounds PZ and NR for Experiment 6 and the compounds PR and QS for Experiment 5. Both experiments showed that the predictive cues received more attention than the uncertain cues. Therefore, it can be concluded that the nature of the uncertain compounds in the test stage was not the cause of the observed differences in learning effects from previous experiments.

Experiments 2 and 4 showed that uncertain cues received more attention than predictive cues (learned uncertainty), but Experiments 5 and 6 showed that predictive cues received more attention than uncertain cues (learned predictiveness). Experiments 4, 5 and 6 were an attempt to find the crucial factor that caused these differences in learning effects. The finding of Experiment 4 excluded the factor of stimuli difference, and the findings of Experiments 5 and 6 excluded the factor of the nature of the uncertain compounds in the test stage. Therefore, the complexity of the training procedure seems to be the factor that drives attention to be paid to either predictive cues or uncertain cues. When the task was relatively less complicated (eight certain compounds and four uncertain compounds in Experiments 3 and 4) uncertain cues received more attention than predictive cues. Conversely, when the task was relatively difficult (eight certain compounds and eight uncertain compounds in Experiments 5 and 6) predictive cues received more attention than uncertain cues.

Experiment 7: uncertain vs. predictive/low difficulty vs high difficulty

Introduction:

So far, the uncertainty effect (Experiments 2 and 4) and the predictiveness effect (Experiments 5 and 6) were observed under different degrees of task difficulty. Under the simple training procedure (Experiments 2 and 4: four uncertain compounds, eight certain compounds), uncertain cues received more attention than predictive cues. Conversely, under the complex training procedure (Experiments 5 and 6: eight uncertain compounds, eight certain compounds), predictive cues received more attention than uncertain cues. The uncertainty effect and the predictiveness effect were observed in different experiments. Therefore, in the current experiment, the concept that the complexity of the training procedure can determine which of the two learning effects is observed was validated by running both of the experimental designs from Experiments 2 and 5 within the same experiment. There were two groups in Experiment 7: the low difficulty group and the high difficulty group. For the low difficulty group, the experimental design was the same as in Experiment 2 in which there were only four uncertain compounds in the training stage. For the high difficulty group, there were eight uncertain compounds in the training stage and the procedure was exactly the same as in Experiment 5. The primary aim of this experiment was to test whether both learning effects can be obtained within the same experiment. This will also provide an important replication of the observed effects and will allow direct statistical comparisons between the learning effects. If the uncertainty effect is found in the low difficulty group and the predictiveness effect is found in the high difficulty group, then it can be concluded that the complexity of the training procedure determines whether the learned predictiveness or the learned uncertainty effect is seen regardless of whether the comparison is between experiments or within the same experiment. Secondly, the uncertainty effect is rarely found in human literatures. If the result of low difficulty group still showed the uncertainty effect, it could provide robust evidence for the learned uncertainty. Finally, the sample size of this experiment (twenty-four participants for each group) differed from the previous experiment (thirty-two participants for both Experiment 2 and Experiment 5). It could also test whether the task difficulty can drive different learning effects in a smaller sample size.

Participants:

In the low difficulty group, twenty-four people (7 males and 17 females) participated in the experiment. The age range was 18-31 (mean: 23.9, SD: 6.0). In the high difficulty group there were twenty-four participants (9 males and 13 females) and the age range was 19-35 (mean = 22.8, SD = 4.1) All participants had normal or corrected to normal vision. Durham University Psychology students received participant pool credit, and other participants were compensated for their time at a rate of £10/hour.

Apparatus

All the details are the same as for the previous experiments.

Procedures:

Participants (N=48) were randomly assigned to either the low difficulty group (N=24) or the high difficulty group (N=24). The procedure of the low difficulty group was the same as the procedure of Experiment 2 (four uncertain compounds for stage 1). The procedure of the high difficulty group was the same as the procedure of Experiment 5 (eight uncertain compounds for stage 1). All the details of the low difficulty group were identical to Experiment 2, and all the details of the high difficulty group were the same as Experiment 5.

Results:

Stage 1: Accuracy increased across blocks for certain compounds for both groups, but this improvement was not seen for uncertain compounds (see Figure 20). A three-way ANOVA of block (1-4) by group (low difficulty and high difficulty) by certainty (certain and uncertain) on accuracy showed that there was a significant main effect of certainty [$F(1,46) = 87.59$, $p < 0.001$, $\eta_p^2 = .66$, 90% CI [.51, .74], power = 1.00], which interacted with block [$F(3,138) = 16.65$, $p < 0.001$, $\eta_p^2 = .27$, 90% CI [.15, .35], power = 1.00]. Block also interacted with group [$F(3,138) = 2.79$, $p = 0.043$, $\eta_p^2 = .06$, 90% CI [.00, .11], power = .68]. However, there was no significant interaction between certainty and group [$F(1,46) = 2.21$, $p = 0.14$, $\eta_p^2 = .05$, 90% CI [.00, .17], power = .32], and no significant three-way interaction among certainty, block and group [$F < 1$, $p = 0.53$], which suggested that any difference between the two groups over the course of training was not specific to the certain condition.

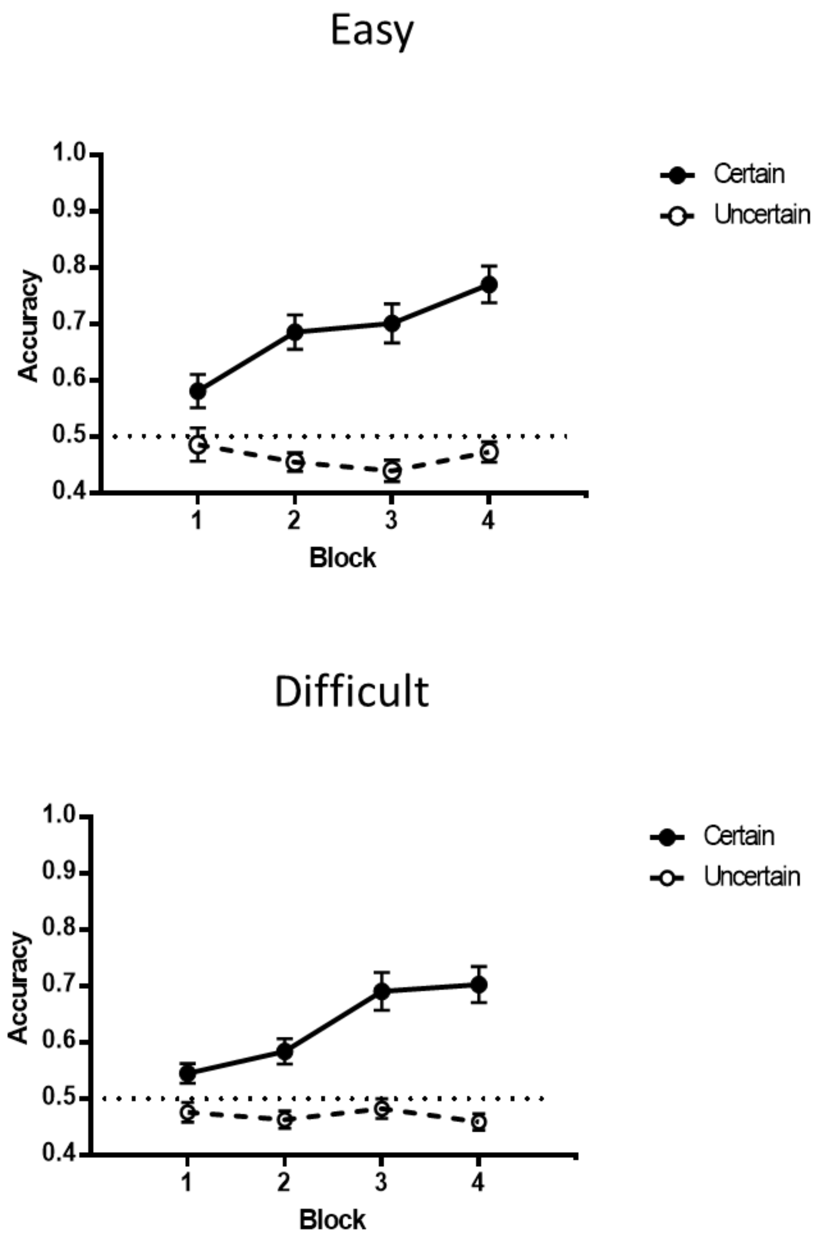

Figure 20. Accuracy across four blocks of stage 1 for Easy group (upper panel) and Difficult group (lower panel). Error bars indicate ± the standard error of mean. The dash line on the left panel indicates chance level (0.5)

Stage 2: Accuracy increased across blocks for all cue compounds for both groups (see Figure 21). A three-way ANOVA of block (1-4) by group (low difficulty and high difficulty) by condition (recombined and control) on accuracy showed that there was a significant main effect of block [$F(3,69) = 52.59$, $p < 0.001$, $\eta_p^2 = .53$, 90% CI [.43, .60], power = 1.00], but no other significant effects or interactions (F values < 1.6, p values > 0.2).

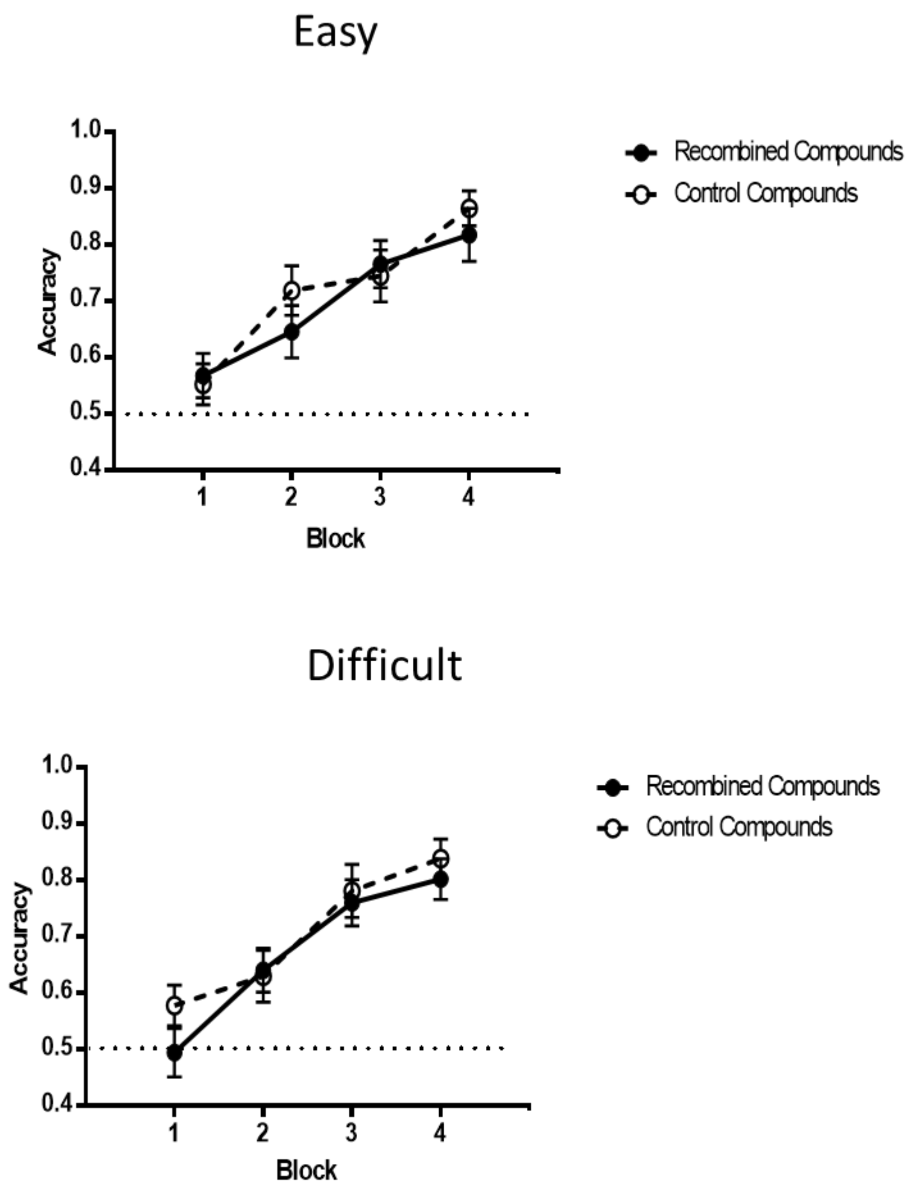

Figure 21. Accuracy across four blocks of stage 2 for Easy group (upper panel) and Difficult group (lower panel). Error bars indicate ± the standard error of mean. The dash line on the left panel indicates chance level (0.5)

Test stage: The ratings for all of the test stage compounds for both groups can be seen in Figure 22. The mean ratings of cue compounds for both groups paired with outcome 4 (BD and QS) were higher than the mean ratings of cue compounds paired with outcome 3 (AC and PR), suggesting that participants learned the causal relationships of the cue-outcome associations. Importantly, a three-way ANOVA of condition (predictive cues vs. uncertain cues) by outcome (outcome 3 vs. outcome 4) by group (low difficulty vs. high difficulty) showed a significant three-way interaction [$F(1,46) = 24.50$, $p < 0.001$, $\eta_p^2 = .35$, 90% CI [.16, .49], power = 1.00], suggesting that the key factor to cause the difference between the results of Experiments 2 and 5 was the complexity of the training procedure. There was also a significant main effect of outcome [$F(1,46) = 56.25$, $p < 0.001$, $\eta_p^2 = .55$, 90% CI [.38, .66], power = 1.00], but no significant main effect of condition [$F(1,46) = 2.73$, $p = 0.11$, $\eta_p^2 = .06$, 90% CI [.00, .19], power = .32]. The interaction between condition and outcome was not significant [$F < 1$, $p = 0.70$], and all other effects and interactions were not significant [F values < 2.1, p values > 0.16]. For the low difficulty group, the difference between the ratings for the predictive cues was smaller than for the uncertain cues, suggesting that predictive cues received less attention than uncertain cues.

The significant three-way interaction was further investigated by carrying out separate two-way ANOVAs for each group. For the low difficulty group, the difference between the ratings for the predictive cues was smaller than for the uncertain cues, suggesting that predictive cues received less attention than uncertain cues. There was a significant main effect of outcome [$F(1,23) = 14.15$, $p = 0.001$, $\eta_p^2 = .38$, 90% CI [.04, .46], power = .96], but no significant main effect of condition [$F(1,23) = 1.47$, $p = 0.24$, $\eta_p^2 = .06$, 90% CI [.00, .25], power = .23]. The interaction between condition and outcome was significant [$F(1,23) = 8.21$, $p = 0.009$, $\eta_p^2 = .26$, 90% CI [.04, .46], power = .82], suggesting that the difference between uncertain cues was larger than the difference between predictive cues. Simple main effects analysis showed that the rating for compound AC was significantly higher than the rating for compound PR [$F(1,23) = 8.97$, $p = 0.006$, $\eta_p^2 = .28$, 90% CI [.05, .48], power = .85] and the rating for BD was not significantly different from the rating for QS [$F(1,23) = 1.98$, $p = 0.17$, $\eta_p^2 = .08$, 90% CI [.00, .27], power = .29]. The rating for AC was not significantly different from the rating for BD [$F(1,23) = 1.23$, $p = 0.28$, $\eta_p^2 = .05$, 90% CI [.00, .23], power = .20]. However, the rating for QS was significantly higher than the rating for PR [$F(1,23) = 29.66$, $p < 0.001$, , $\eta_p^2 = .56$, 90% CI [.30, .69], power = 1.00].

For the high difficulty group, the difference between the ratings for the predictive cues was larger than for the uncertain cues, indicating that predictive cues received more attention than uncertain cues. There was a significant main effect of outcome [$F(1,23) = 57.32$, $p < 0.001$, $\eta_p^2 = .71$, 90% CI [.51, .80], power = 1.00], but no significant main effect of condition [$F(1,23) = 1.27$, $p = 0.64$, $\eta_p^2 = .05$, 90% CI [.00, .24], power = .20]. The interaction between condition and outcome was significant [$F(1,23) = 19.47$, $p < 0.001$, $\eta_p^2 = .46$, 90% CI [.19, .62], power = .99], suggesting that the difference between uncertain cues was larger than the difference between predictive cues. Simple main effects analysis showed that the rating for compound AC was significantly higher than the rating for compound PR [$F(1,23) = 4.15$, $p = 0.053$, $\eta_p^2 = .15$, 90% CI [.00, .36], power = .53] and the rating for BD was not significantly different from the rating for QS [$F(1,23) = 27.48$, $p < 0.001$, $\eta_p^2 = .54$, 90% CI [.28, .68], power = .99]. The rating for AC was not significantly different from the rating for BD [$F(1,23) = 77.23$, $p < 0.001$, $\eta_p^2 = .77$, 90% CI [.59, .84], power = 1.00]. However, the rating for QS was significantly higher than the rating for PR [$F(1,23) = 9.71$, $p < 0.001$, $\eta_p^2 = .30$, 90% CI [.06, .49], power = .88].

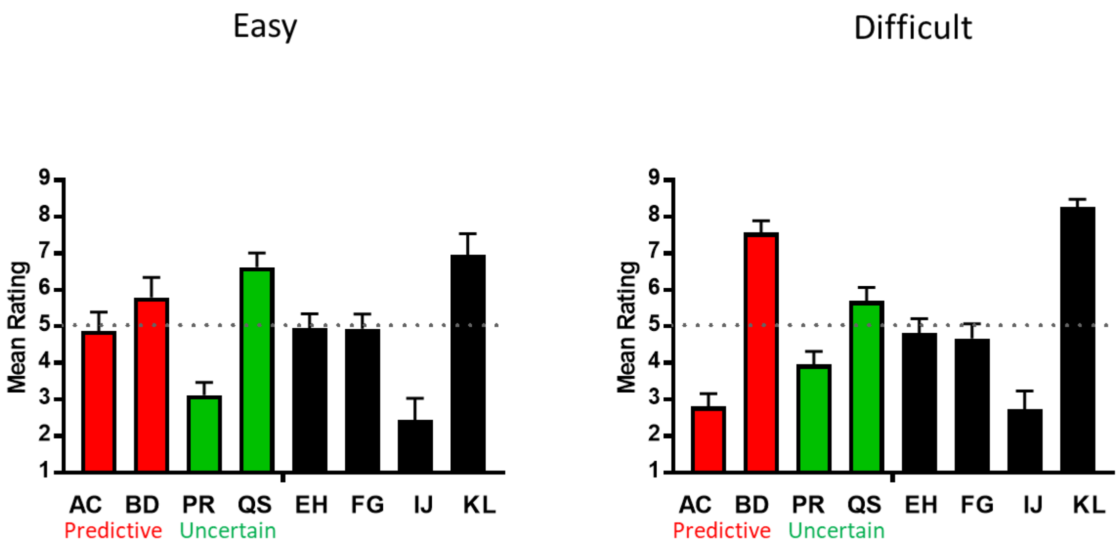

Figure 22. The rating scores of each compound in test stage. The Y axis is the mean rating score. Error bars stand for standard error of mean. Left panel is for the easy group; while right panel is for the hard group.

Discussion:

In Experiment 7 there were two groups: the low difficulty group (four uncertain compounds in stage 1) and the high difficulty group (eight uncertain compounds in stage 1). The low difficulty group replicated the findings of Experiment 2, suggesting that predictive cues received less attention than uncertain cues. The high difficulty group replicated the results of Experiment 5, indicating that attention paid to predictive cues was greater than attention paid to uncertain cues. These opposite results indicate that the key factor that causes this difference is the complexity of the stage 1 training procedure, determined by the number of uncertain compounds present during training.

Experiment 8 : Uncertain vs. Predictive – four extra predictive compounds

Introduction:

The previous experiments (from Experiment 2 to Experiment 7) demonstrated that the addition of four extra uncertain compounds during training caused a change in attention from uncertain to predictive cues. When the task difficulty (the complexity of the training procedure) was relatively low (Experiments 2, 4, and the low difficulty group of Experiment 7), attention paid to uncertain cues was greater than predictive cues. However, predictive cues received more attention than uncertain cues when the difficulty of the training procedure was relatively high (Experiments 5, 6, and the high difficulty group of Experiment 7). Experiment 7 confirmed that the task difficulty influences the effects of learned predictiveness and learned uncertainty. In other words, under relatively simple training procedure (4 uncertain compounds and 12 certain compounds in stage 1), attention paid to uncertain cues was higher than attention paid to predictive cues; while, under relatively difficult training procedure (8 uncertain compounds and 8 certain compounds in stage 1), the pattern was reverse. It is not clear, however, exactly what constitutes a high difficulty task. It seems to that the addition of four uncertain compounds switched attention from uncertain to predictive cues, but it is not known if the difficulty increase was due to the addition of uncertain cues or an increase in the required memory load for the task. This question will be addressed in Experiment 8.

In the low difficulty training procedure, there were eight certain compounds and four uncertain compounds, while in the high difficultly training procedure there were eight certain compounds and eight uncertain compounds. The difference between these two procedures is that there were an extra four uncertain compounds in the high difficulty training procedure. However, the four extra uncertain compounds could also have increased memory load in the high difficulty task due to an increase in the number of cues. Therefore, whether the task difficulty is related to the number of uncertain cues specifically, or the number of cues in general, will be explored in Experiment 8. In order to do that, the four extra uncertain compounds in the high difficulty training group will be replaced by four certain compounds. In this way, twelve certain compounds and four uncertain compounds will be presented to participants in Experiment 8. If the effects of learned predictiveness and learned uncertainty are dependent on the number of uncertain compounds, the results of Experiment 8 should be similar to the results of Experiments 2 and 4 showing the uncertainty effect, because the number of uncertain compounds is the same (four uncertain compounds). If the learning effects are determined by the memory load, then the results of Experiment 8 should be similar to those of Experiments 5 and 6 showing the predictiveness effect, as there are 16 cue compounds during training. However, if the learning effects showed that the amount of attention paid to predictive cues was similar to the amount of attention paid to uncertain cues, it might potentially suggest that the learned predictiveness and the learned uncertainty depend on the task difficulty, as the complexity of training procedure of Experiment 8 (4 uncertain compounds, 12 certain compounds) is between the relative simple training procedure (4 uncertain compounds, 8 certain compounds) and the relatively difficult training procedure (8 uncertain compounds, 8 certain compounds).

Participants:

Twenty-four people (7 males and 17 females) participated in the experiment. The age range was 19-26 (mean = 21.1, SD = 2.1). All participants had normal or corrected to normal vision. Durham University Psychology students received participant pool credit, and other participants were compensated for their time at a rate of £10/hour.

Apparatus

All the details are the same as for previous experiments.

Procedures:

In stage 1, there were twelve certain compounds (AV→O1, BV→O2, AW→O1, BW→O2, CX→O2, DX→O1 CY→O1, DY→O2, ZM→O1, ZO→O2, NO→O1, NM→O2) and 4 uncertain compounds (PQ→O1/O2, PS→O1/O2, RS→O1/O2, RQ→O1/O2), see Table 8. Stage 2 and the test stage were exactly the same as for Experiment 6 and the high difficulty group of Experiment 7. All other details were the same as for previously experiments.

Table 8. Design of Experiment 8. Letters represent cues (presented in compound) and numbers represent outcomes. During the test stage, participants were asked to rate the expected likelihood of outcome 3 or outcome 4 given the presented cue compound.

Stage 1	Stage 2	Test
AV → O1	AP → O3	AC → O3/O4?
BV → O2	BQ → O4	BD → O3/O4?
AW → O1	CR → O3	PR → O3/O4?
BW → O2	DS → O4	QS → O3/O4?
CX → O2	EF → O3	EH → O3/O4?
DX → O1	GH→ O4	FG → O3/O4?
CY → O2	IJ → O3	IJ → O3/O4?
DY → O1	KL → O4	KL → O3/O4?
PQ → O1/O2		
PS → O1/O2		
RS → O1/O2		
RQ → O1/O2		
ZM → O1		
ZO → O1		
NO → O2		
NM → O2		

Predictive cues
Irrelevant cues
Uncertain cues

Results:

Stage 1: Accuracy increased across blocks for certain compounds, but this was not the case for uncertain compounds (see Figure 23). A two-way ANOVA of block (1-4) by certainty (certain and uncertain) on accuracy showed that there was a significant main effect of certainty [F(1,23) = 83.31, p < 0.001, η_p^2 = .78, 90% CI [.61, .85], power = 1.00], and a significant main effect of block [F(3,69) = 4.85, p = 0.04, η_p^2 = .17, 90% CI [.04, .28], power = .91] and the interaction between these factors was significant [F(3,69) = 8.78, p < 0.001, η_p^2 = .28, 90% CI [.11, .38], power = 1.00].

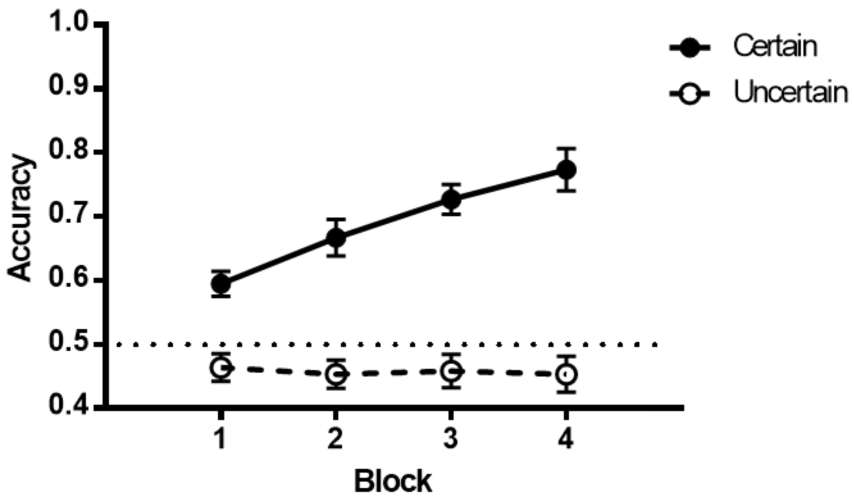

Figure 23. Accuracy across four blocks of stage 1. Error bars indicate ± the standard error of mean. The dash line on the left panel indicates chance level (0.5)

Stage 2: Accuracy increase over blocks for all cue compounds (see Figure 24). A two-way ANOVA of block (1-4) by trial-type (recombined compounds and control compounds) on accuracy showed that there was a significant main effect of block [$F(3,69) = 38.41$, $p < 0.001$, $\eta_p^2 = .63$, 90% CI [.49, .69], power = 1.00], but no significant main effect of trial-type [$F(1,23) = 1.18$, $p = 0.29$, $\eta_p^2 = .05$, 90% CI [.00, .23], power = 1.00]. There was no significant interaction between these factors [$F(3,69) = 1.91$, $p =0.14$, $\eta_p^2 = .08$, 90% CI [.00, .16], power = .50].

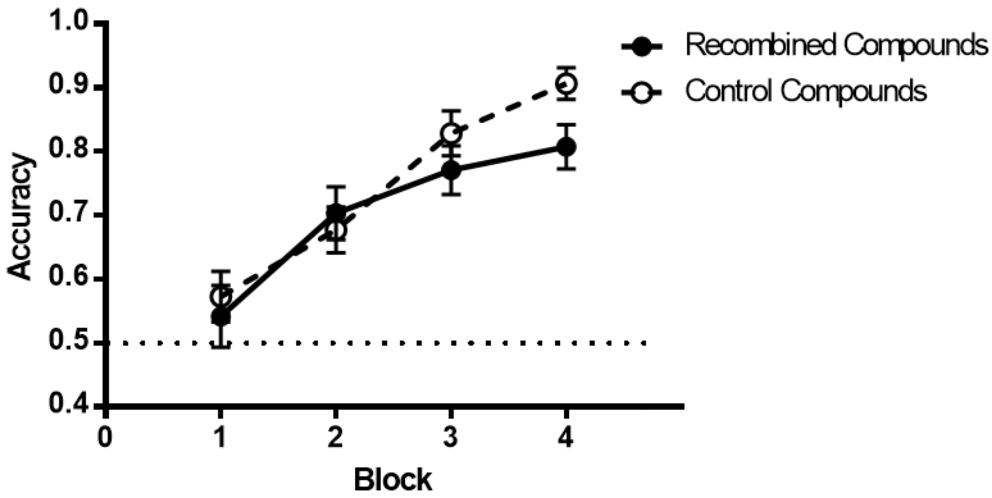

Figure 24. Accuracy in four blocks in stage 2. Error bars indicate ± the standard error of mean. And the dash line on the left panel indicates chance level (0.5)

Test Stage: The ratings for each compound in the test stage are shown in Figure 25. The ratings for compounds consisting of cues paired with outcome 4 (BD and QS) were higher than for those paired with outcome 3 (AC and PR), suggesting that participants had learnt the cue-outcome associations. The difference in ratings between predictive cues was similar to the difference between uncertain cues. A two-way ANOVA of condition (predictive cues AC and BD vs. uncertain cues PR and QS) by outcome (outcome 3 for compounds AC and PR vs. outcome 4 for compounds BD and QS) showed that there was a significant main effect of outcome [$F(1,23) = 46.32$, $p < 0.001$, $\eta_p^2 = .67$, 90% CI [.44, .77], power = 1.00], but no significant main effect of condition [$F(1,23) = 2.63$, $p = 0.32$, $\eta_p^2 = .10$, 90% CI [.00, .30], power = .37]. There was no significant interaction between condition and outcome [$F < 1$, $p = 0.71$] demonstrating that the difference between compounds AC and BD was similar to the difference between PR and QS. This suggests that predictive cues received a similar level of attention as uncertain cues. One-way ANOVA was carried out to test is there any difference between compounds EH and FG and difference between compound IJ and KL. There was no significant difference between compounds EH and FG [$F(1,23) = 1.05$, $p = 0.31$, $\eta_p^2 = .04$, 90% CI [.00, .22], power = .18], and two one-sample t-tests showed that the ratings for both compounds (EH and FG) were not significantly different from a ratings of 5 [ts <1.2, ps > 0.25]. However, there was a significant difference between compounds IJ and KL [$F(1,23) = 26.45$, $p < 0.001$, $\eta_p^2 = .54$, 90% CI [.27, .67], power = .99].

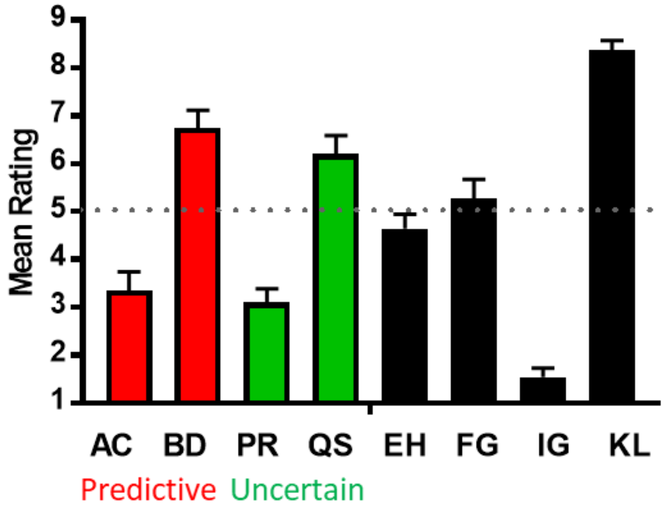

Figure 25. The ratings for each compound in the test stage. The y-axis is the mean rating, with 1 indicating a strong link with outcome 3 and 9 indicating a strong link with outcome 4. The dashed line represents a rating of 5, which indicates that the compound is linked equally with outcomes 3 and 4. Error bars indicate the standard error of the mean.

Correlation Analysis:

From Experiment 2 to Experiment 8, the main finding was that the learned predictiveness and learned uncertainty depends on the task difficulty. So far, there were three levels of the task difficulty: low difficulty (Experiment 2 and the easy group of Experiment 7), middle difficulty (Experiment 8) and high difficulty (Experiment 5, 6 and the difficult group of Experiment 7). Under the simple training procedure, Experiment 2 and the easy group of Experiment 7 showed the learned predictiveness effect; Under the middle degree of difficulty training procedure, Experiment 8 showed that predictive cues receive similar level of attention than uncertain cues; while, under the high difficulty of training procedure, the learned uncertainty effects were observed. Here, I correlated the learning effects (rating scores in the test phase) and task difficulty (3 levels: low difficulty, middle difficulty, high difficulty) to test whether different learning effects (learned predictiveness and learned uncertainty) are modulated by the task difficulty. The learning effect was defined by the rating score difference between predictive compounds (BD-AC) minus the difference between uncertain compounds (QS-PR). In Figure 26, it showed that the learning effects is correlated with task difficulty (R^2 =0.99, p = 0.01), suggesting that the learned predictiveness and learned uncertainty may rely on task difficulty.

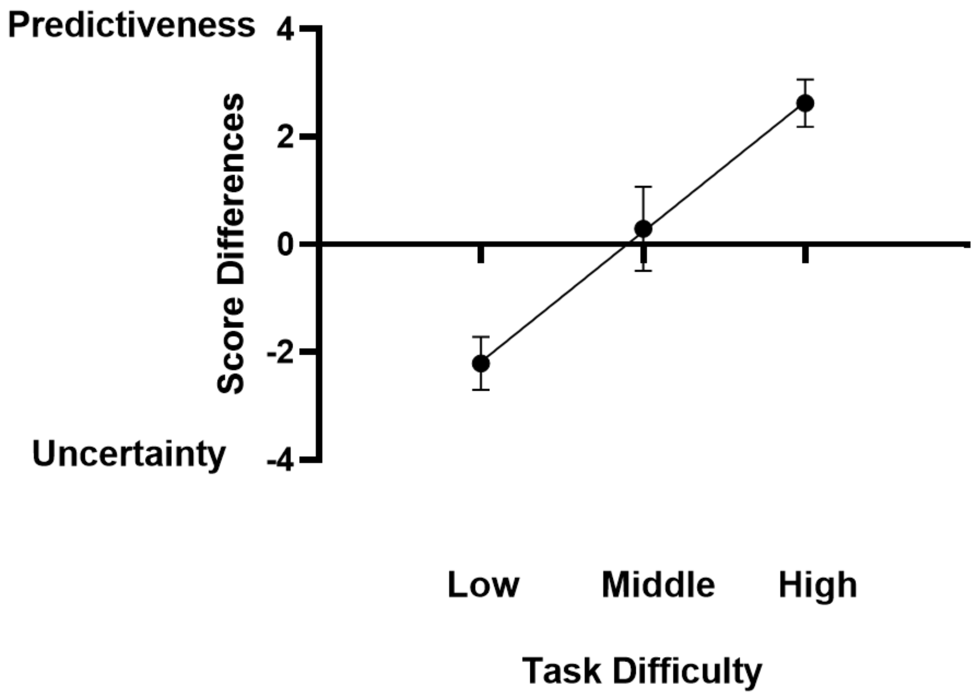

Figure 26. The correlation between learning effects and different levels of task difficulty was shown. The Y axis is the rating score difference; The x axis is the task difficulty. Error bars indicate ± the standard error of mean.

Discussion:

The results of Experiment 8 showed that predictive cues received similar levels of attention to uncertain cues. Unlike Experiment 2 (which contained eight certain compounds and four uncertain compounds) and Experiment 5 (which used eight certain compounds and eight uncertain compounds), this Experiment 8 included twelve certain compounds and four uncertain compounds. In terms of memory load, this experiment should be similar to Experiment 5, because both experiments contain 16 cue compounds in stage 1. However, Experiment 8 failed to replicate the results of Experiment 5 (that predictive cues received more attention than uncertain cues), but also failed to show the opposite result. This may suggest that the number of uncertain compounds and the number of cues could both be important factors in determining the learning effect observed. In terms of the complexity of the training procedure, Experiment 2 was relatively easy, as there were eight certain compounds and four uncertain compounds. Experiment 5 was relatively hard, as there were eight certain compounds and eight uncertain compounds. However, it is likely that the task difficulty of Experiment 8 was somewhere in-between Experiments 2 and 5 (twelve certain compounds and four uncertain compounds). In terms of the findings of the three procedures, the learned predictiveness effect was observed in the more difficult training procedures (Experiments 5 and 6), the learned uncertainty effect was obtained in the relatively easy training procedures (Experiments 2 and 4), and similar levels of learning for predictive cues and uncertain cues was observed when the difficulty of the training procedure was intermediate. Therefore, to some extent, the complexity of the training procedure can drive the different learning effects. In other words, the manner in which attention is allocated can be shaped by the task difficulty.

General Discussion:

In general, the purpose of this chapter was to determine the crucial factor that caused the different learning effects observed in other experiments: learned uncertainty (Experiments 2 and 3) and learned predictiveness (e.g., Livesey et al., 2011). A series of experiments have been conducted to test three differences between the experiments of Chapter 2 and those of Livesey et al.: the stimuli and the cover story (Experiment 4), the complexity of the training procedure (Experiment 5), and the test procedure of uncertain compounds (Experiment 6). The main findings of this chapter (Experiments 4 - 7) showed that participants paid more attention to predictive cues than uncertain cues when they were presented with complex information during the training stage (eight uncertain compounds). This pattern of attention was reversed when the training procedure was relatively simple (four uncertain compounds). In other words, task difficulty altered the way participants paid attention to cues. This effect was confirmed by Experiment 7, which provided a clear replication of the effects of Experiments 2 and 5. One possibility to explain the data is that different attentional strategies were used according to different training procedures. The relatively easy training procedure allowed participants to utilize the attentional exploration strategy in order to reduce uncertainty. In other words, they allocated more attention to uncertain cues to try to determine the relationship between the uncertain cues and the outcomes. On the contrary, when the difficulty of the training procedure was increased, participants used the attentional exploitation strategy to focus on the predictive cues in order to be able to complete the task, perhaps because they did not have enough cognitive resources to try to solve the association between the uncertain cues and the outcomes (Beesley et al., 2015).

From Experiments 2 to 8, the different degrees of task difficulty were formed by the different combinations of the number of uncertain compounds and certain compounds during stage 1 training. The degree of task difficulty of Experiment 8 (twelve certain compounds, four uncertain compounds) is likely to be in-between the difficulties of Experiment 2 (eight certain compounds, four uncertain compounds) and Experiment 6 (eight certain compounds, eight uncertain compounds). The results showed that the learning effect was also in-between learned predictiveness and learned uncertainty in that attention appeared to be paid similarly to both types of cue. Another way to explain the results of Experiment 8 is that the lack of difference between predictive cues and uncertain cues might be simply due to a null finding, and nothing to do with the complexity of the training procedure. However, given all the experiments conducted in this chapter, it is perhaps more sensible to conclude that the complexity of the training procedure is the key factor that causes different learning effects. In this way, the task difficulty is not simply due to either the number of cue compounds or the memory load, but is instead dependent on the combination of the number of uncertain compounds and the number of cues.

Chapter 4:

Introduction

The experiments reported in this chapter were designed to test whether associability is governed by summed or individual prediction error. The results from Chapter 2 showed that uncertain cues (high summed prediction error) received more attention than predictive cues (low summed prediction error) and irrelevant cues (low summed prediction error). The results suggested that attention can be controlled by the size of prediction error on a summed error term, supporting Pearce-Hall attentional processes. However, the reason why uncertain cues received more attention than irrelevant cues could be explained by learning based simply on the predictive cues (see the general discussion in chapter 2). By testing the changes of attention to biconditional cues, this chapter will further investigate the relationship between associability and different forms of prediction error. Biconditional cues (e.g., PQ-1, PS-2, RS-1, RQ-2) have low summed prediction error but high individual prediction error, as the compounds (PQ, PS, RS, RQ) consistently lead to specific outcomes, but, individually, biconditional cues (P,Q,R,S) are partial reinforced. Therefore biconditional discriminations offer the opportunity to test two forms of prediction error (high individual prediction error, low summed prediction error), which is different from uncertain compounds (high summed prediction error & high individual prediction error) and predictive compounds (low summed prediction error & low individual prediction error). Thus, whether the summed prediction error or individual prediction error modulate the associability in associative learning can be further investigated by comparing predictive cues, uncertain cues and biconditional cues.

Some studies have shown that the biconditional cues receive more attention than irrelevant cues (George & Pearce, 1999; Kruschke, 1996). However, Livesey, et al. (2011) failed to find attention paid to biconditional cues was greater than attention paid to irrelevant cues. Instead, they found biconditional cues received similar level of attention to irrelevant cues. However, for the comparison between irrelevant cues and biconditional cues, both types of cues have high individual prediction error but low summed prediction error. Therefore, for the purposes of the focus of this thesis, this comparison would not be expected to help to determine which form of prediction error influences associability. Instead, in Experiment 9, reported in the current chapter, the comparison between uncertain cues and biconditional cues was made. If associability is driven by summed prediction error, uncertain cues should receive more attention than biconditional cues, as the summed prediction error of uncertain compounds is higher than biconditional compounds. In contrast, if associability is controlled by individual prediction error, then attention paid to biconditional cues should be similar to uncertain cues, as they both have high individual prediction error term. Similarly, in Experiment 10, the comparison between biconditional cues and predictive cues was made. If learning was influenced by the summed prediction error, attention paid to biconditional cues should receive more attention, since the summed prediction error of biconditional cues is higher than predictive cues; while if learning was driven by the individual prediction error, biconditional cues should receive similar degree of attention to predictive cues, because both biconditional cues and predictive cues have low individual prediction error. By running Experiment 9 and Experiment 10, investigating the role of biconditional cues could possibly solve the debate between learning theories:

whether the individual prediction error or the summed prediction error can drive the associability.

Another goal of this chapter is to test whether task difficulty can determine how much attention participants pay to a cue when the comparison includes biconditional cues (Experiment 11). In Chapter 3, task difficulty was identified as a potential crucial factor to drive participants' attention. When the difficulty of the training procedure was relatively easy (Experiment 2, Experiment 4 and the simple version of Experiment 7), attention was allocated to uncertain cues; conversely, attention was paid to predictive cues when task difficulty of was relatively high (Experiment 5 and Experiment 6). Experiment 9 compared biconditional cues to uncertain cues, with the results showing that uncertain cues received a similar level of attention to biconditional cues. In Experiment 10, the comparison between biconditional cues and predictive cues showed that biconditional cues received more attention than predictive cues. Taken together, Experiment 9 and Experiment 10 might potentially suggest that biconditional cues functioned in a similar way to uncertain cues. If the role of biconditional cues functioned in a manner similar to that of uncertain cues, it could be anticipated that when the difficulty of training procedure is high, predictive cues receive more attention than biconditional cues. In contrast, attention paid to biconditional cues should be higher than predictive cues when the difficulty of training procedure is low. The difficulty of training procedure of Experiment 10 (4 biconditional compounds & 8 certain compounds) was similar to Experiment 2 and Experiment 4 (4 uncertain compounds & 8 certain compounds), given biconditional cues play similar role to uncertain cues. Thus, in Experiment 11, the task difficulty increased by adding four extra uncertain compounds. In total, there were 8 certain compounds composed of predictive cues and irrelevant cues, 4 uncertain compounds and 4 biconditional compounds, which is similar to Experiment 5 and Experiment 6 (8 certain compounds & 8 uncertain compounds). It could be anticipated that attention should be greater to predictive cues than biconditional cues.

Methods

Apparatus

All the details are the same as Chapter 2

Behavioural analysis

All the details are the same as Chapter 2

Experiment 9

Introduction:

Experiment 9 compared biconditional cues to uncertain cues in order to test whether the associability was modulated by summed prediction error or individual prediction error. The summed prediction error of biconditional cues is low, but the summed prediction error of uncertain cues is high. If summed prediction error determines the associability, it could be anticipated uncertain cues received more attention than biconditional cues; while if individual prediction error determines how much attention participants pay to a cue, the results should show biconditional cues receive a similar level of attention to uncertain cues, as the individual prediction errors of biconditional cues and uncertain cues are both low.

Participants:

Thirty-two people (6 males and 26 females) participated in the experiment. The age range was 18-29 (mean: 21.7, SD: 2.5). All participants had normal or corrected to normal vision. Durham University Psychology students received participant pool credit and other participants were compensated for their time at a rate of £10/hour.

Procedures:

In stage 1, there were two types of cue compounds: uncertain cue compounds and biconditional cue compounds (see Table 9). Uncertain cue compounds (PQ → 1/2, PS → 1/2, RS → 1/2, RQ → 1/2) were the same with previous Experiments 2-8.

The biconditional cue compounds (AB →1, CD → 1, AD →2, CB → 2) were predictive of specific outcomes, but individually they led to different outcomes, see Table 7. There were eight trial types in stage 1 (AB →1, CD → 1, AD →2, CB → 2, PQ → 1/2, PS → 1/2, RS → 1/2, RQ → 1/2). Each trial type was repeated 16 times. In total, there were 128 trials. In stage 2, the procedure was similar to stage 1. The major difference between stages was the different outcomes (outcome 1 & 2 for stage 1; outcome 3 & 4 for stage 2). Additionally, in Stage 2, each cue was predictive of the outcome (AP→O3, BQ→O4, CR→O3, DS→O4, EF→O3, GH→O4, IJ→O3, KL→O4). Participants received the eight trial types in which pairs of flags reliably led to either outcome 3 or 4 (See Table 6, stage 2). The first four of eight trial types consisted of pairs of flags that included one biconditional cue (A,B,C,D) and one uncertain cue (P,Q,R,S) from stage 1 (recombined cues: AP→O3, BQ→O4, CR→O3, DS→O4). For the remaining trial types, new flags that were previously not experienced in stage 1 were used (EF→O3, GH→O4, IJ→O3, KL→O4), which is the same as in previous experiments. There were 64 trials in stage 2, in which each trial type repeated 8 times. In the test phase, participants were asked to rate how likely the presented compounds (AC, BD, PR, QS, EH, FG, IJ, KL) led to outcome 3 or outcome 4. There were eight trial types. Half of the trial types (compounds EH,FG,IJ,KL) were utilized to test whether participants could use this rating scale properly. The other half of tested compounds (AC,BD,PR,QS) were used to examine the learning effect of stage 2 in order to know whether the stage 1 training can influence stage 2 learning. All other details were the same as for previous experiments.

Table 9. Design of Experiment 9. Letters represent cues (presented in compound) and numbers represent outcomes. During the test stage, participants were asked to rate the expected likelihood of outcome 3 or outcome 4 given the presented cue compound.

Stage 1	Stage 2	Test
AB → O1	AP → O3	AC → O3/O4?
CD → O1	BQ → O4	BD → O3/O4?
AD → O2	CR → O3	PR → O3/O4?
BC → O2	DS → O4	QS → O3/O4?
PQ → O1/O2	EF → O3	EH → O3/O4?
PS → O1/O2	GH → O4	FG → O3/O4?
RS → O1/O2	IJ → O3	IJ → O3/O4?
RQ → O1/O2	KL → O4	KL → O3/O4?

Biconditional cues
Uncertain cues

Results

Stage 1: Participants acquired the discrimination over training with performance increasing for the biconditional compounds over blocks, but showed no improvement for the uncertain compounds (see Figure 27). A two-way ANOVA was carried out to test whether block (1-4) and certainty (biconditional & uncertain compound) had an effect on accuracy in stage 1. There was a significant effect of block [$F(3,93) = 6.13$, $p < 0.001$, $\eta_p^2 = .17$, 90% CI [.05, .26], power = .96] and certainty [$F(1,31) = 17.30$, $p < 0.001$, $\eta_p^2 = .36$, 90% CI [.14, .52], power = .99], and a significant interaction between block and certainty [$F(3,93) = 3.70$, $p < 0.014$, $\eta_p^2 = .11$, 90% CI [.01, .19], power = .81].

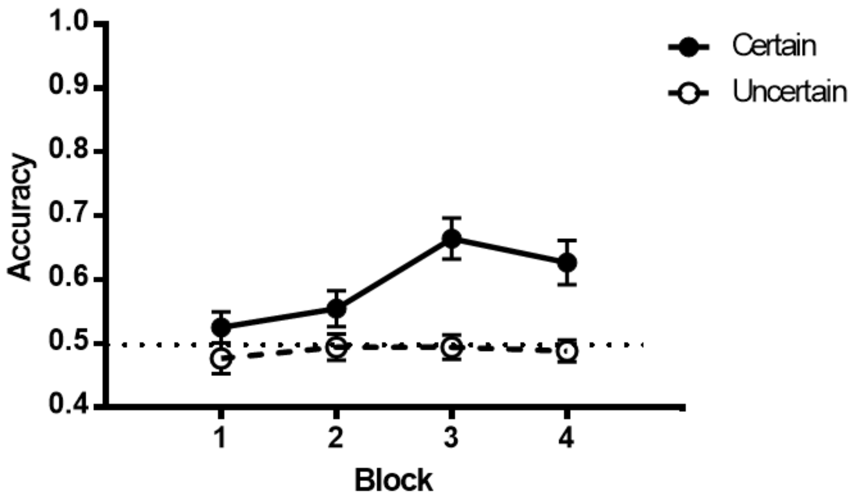

Figure 27. Accuracy across four blocks of stage 1. Error bars indicate ± standard error of mean. The dash line on the left panel indicates chance level (0.5)

Stage 2: Participants acquired the association between cues and outcomes over training for both the recombined compounds and the control compounds (see Figure 28). A two-way ANOVA was applied to analyse whether the two factors (trial-type and block) have an effect on accuracy. It revealed that there was a significant main effect of block [$F(3,93) = 38.87$, $p < 0.001$, $\eta_p^2 = .56$, 90% CI [.43, .63], power = 1.00], and of trial-type [$F(1,31) = 8.72$, $p = 0.006$, $\eta_p^2 = .22$, 90% CI [.04, .40], power = .84], but there was no interaction between factors [$F(3,93) = 2.69$, $p = 0.061$, $\eta_p^2 = .08$, 90% CI [.00, .16], power = .66].

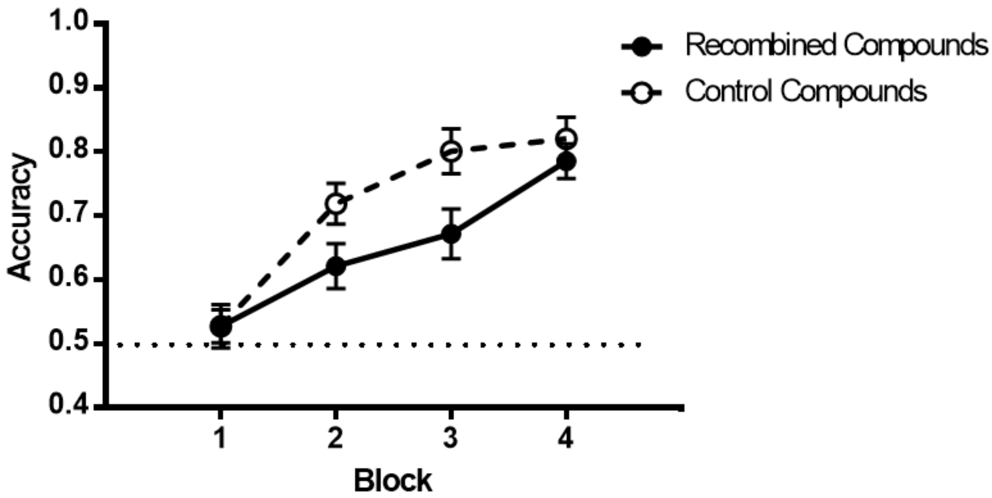

Figure 28. Accuracy in four blocks in stage 2. Error bars stand for ± standard error of mean. And the dash line on the left panel indicates chance level (0.5)

Test phase: The ratings for the test stage are shown in Figure 29. The ratings for compounds consisting of cues paired with outcome 4 (BD and QS) were higher than for those paired with outcome 3 (AC and PR), suggesting that participants had learnt the cue-outcome associations. More importantly, the difference between AC and BD was similar to the difference between PR and QS. A two-way ANOVA (condition: biconditional cues (AC, BD) vs. uncertain cues (PR, QS); outcome: 3 (AC, PR) vs. 4 (BD, QS)) showed that there was a significant main effect of outcome [$F(1,31) = 17.79$, $p < 0.001$, $\eta_p^2 = .36$, 90% CI [.14, .53], power = .99], but no main effect of condition [$F < 1$, $p = 0.48$]. Consistent with the observation made above, there was no interaction between condition and outcome [$F < 1$, $p = 0.44$]. The lack of an interaction might suggest that biconditional cues received a similar level of attention to uncertain cues. Alternatively, it might be that the results simply failed to reveal the effect. There was no difference between compound EH and compound FG [$F(1,31) = 0.07$, $p = 0.94$, $\eta_p^2 = .01$, 90% CI [.00, .08], power = .06], and two one-sample t-tests showed that the scores of both compounds (EH and FG) were not significantly different from chance [$t <0.5$, $p > 0.70$]. However, there was a significant difference between compound IJ and compound KL [$F(1,31) = 6.99$, $p < 0.001$, $\eta_p^2 = .18$, 90% CI [.02, .37], power = .75].

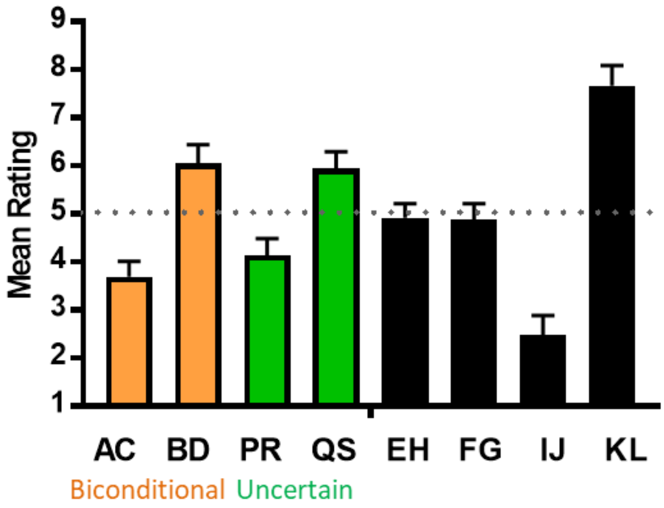

Figure 29. The rating scores of each compound in test phase. The Y axis is the mean rating score. Error bars indicate standard error of mean.

Discussion:

In the current experiment, the main finding was that there was no rating difference between biconditional cues and uncertain cues. There are two ways to interpret this result: firstly, attention paid to biconditional cues was similar to attention paid to uncertain cues. Secondly, this experiment failed to reveal any effect. If the first explanation is true, it potentially suggests that biconditional cues and uncertain cues play similar role in this associative learning. The main purpose of this experiment was to test whether the associability is governed by summed prediction error or individual prediction error. To do that, biconditional cues were applied to this experiment. The compounds (PQ→1, PS→2, RS→1, RQ→2) composed of biconditional cues reliably led to specific outcomes (outcome 1 or outcome 2), so the summed prediction error ($\lambda - \Sigma V$) of the compounds was low; while the individual prediction error ($\lambda - V$) of each biconditional cue in the compound remained high due to each cue (P,Q,R,S) led equally to either outcome 1 or outcome 2. Therefore, the individual prediction error of each biconditional cue in the compound remained high. But the summed prediction error of biconditional compound was low, as the link between biconditional cue compounds and outcomes were consistently reinforced and could be well predicted. In contrast, the individual prediction errors of uncertain cues (PQ → 1/2, PS → 1/2, RS→ 1/2, RQ→1/2) were high due to the partial reinforcement, and the summed prediction errors of uncertain compounds were also high.

If learning is driven by the summed prediction error, attention paid to uncertain cues should be higher than biconditional cues, as the summed prediction error of uncertain cues are higher than the summed prediction error of biconditional cues.

However, the results showed that biconditional cues received similar level of attention to uncertain cues, indicating that summed prediction error might not suitable to explain the results. In contrast, the individual prediction error may explain the data. The individual prediction error of uncertain cues is similar to the individual prediction error of biconditional cues, therefore, they receive similar attention. To test this notion further, Experiment 10 was conducted.

Experiment 10

Introduction:

Experiment 9 might suggest that biconditional cues functioned in a similar manner of uncertain cues. If that is case, biconditional cues should receive more attention than predictive cues as uncertain cues received more attention than predictive cues in Experiment 2. Therefore, in Experiment 10, I compared biconditional cues to predictive cues. Running Experiment 10 can also help us to understand which type of prediction error can determine the associability. If the associability was driven by the summed prediction error term, then biconditional cues and predictive cues should receive similar attention, as the summed prediction errors of biconditional cues and predictive cues were both low. In contrast, if individual prediction error determines how much attention participants paid to a cue, attention paid to biconditional cues should be higher than to predictive cues, because the individual prediction error of biconditional cues was greater than predictive cues. It should be noted that there were only four biconditional compounds in this experiment, which the task difficulty is similar as Experiment 2 (only four uncertain compounds)

Participants:

Twenty-one people (6 males and 15 females) participated in the experiment. The age range was 20-35 (mean: 24.8, SD: 3.3). All participants had normal or corrected to normal vision. Durham University Psychology students received participant pool credit and other participants were compensated for their time at a rate of £10/hour.

Procedures:

In stage 1, there were two types of certain cue compounds: biconditional cue compounds and predictive cue compounds (see Table 10). The predictive compounds always comprised one cue that reliably predicted the outcome, combined with a cue that was irrelevant to the outcome. (AV→1, BV→2, AW→1, BW→2, CX→2, DX→1, CY→2, DY→1). These were the same in previous Experiments 2-8, and the biconditional cue compounds (PQ →1, PS → 2, RS→1, RQ → 2) were the same as in Experiment 9. The procedures of stage 2 were the similar to the previous experiments. Each of the first four compounds of stage 2 were composed of one biconditional cue (P,Q,R,S) and one predictive cue (A,B,C,D). In the test phase, participants were required to rate how likely the test compounds (AC, BD, PR, QS, EH, FG, IJ, KL) led to outcome 3 or outcome 4. All other details were the same to the previous experiments.

Table 10. Design of Experiment 10. Letters represent cues (presented in compound) and numbers represent outcomes. During the test stage, participants were asked to rate the expected likelihood of outcome 3 or outcome 4 given the presented cue compound.

Stage 1	Stage 2	Test
AV → O1	AP → O3	AC → O3/O4?
BV → O2	BQ → O4	BD → O3/O4?
AW → O1	CR → O3	PR → O3/O4?
BW → O2	DS → O4	QS → O3/O4?
CX → O2	EF → O3	EH → O3/O4?
DX → O1	GH→ O4	FG → O3/O4?
CY → O2	IJ → O3	IJ → O3/O4?
DY → O1	KL → O4	KL → O3/O4?
PQ → O1		
PS → O2		
RS → O1		
RQ → O2		

Predictive cues
Irrelevant cues
Biconditional cues

Results:

Stage 1: Mean accuracy in each block was averaged across the predictive cue compounds (AV,BV,AW,BW,CX,DX,CY,DY) and biconditional cue compounds (PQ,PS RQ,RS) as shown in Figure 30. Overall, accuracy in stage 1 for both compounds increased across the blocks (see Figure 29). For accuracy, a repeated measure ANOVA, with block (1-4) and certainty (predictive compounds vs. biconditional compounds) as factors, showed a significant main effect of block [$F(3,60) = 28.80$, $p < 0.001$, $\eta_p^2 = .59$, 90% CI [.43, .67], power = 1.00], but no effect of certainty [$F(1,20) = 2.57$, $p = 0.12$, $\eta_p^2 = .11$, 90% CI [.00, .33], power = .36], and no interaction between them [$F < 1$, $p = 0.43$].

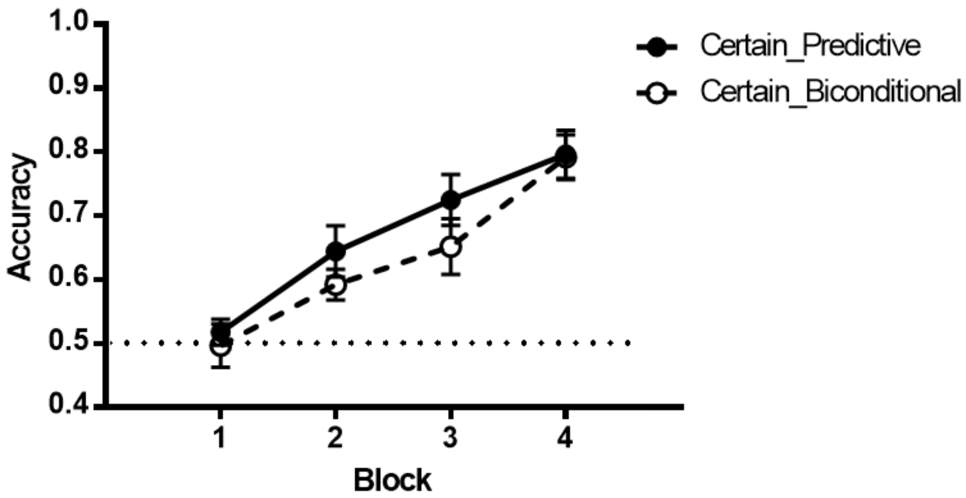

Figure 30. Accuracy across four blocks of stage 1. Error bars indicate ± standard error of mean. The dash line on the left panel indicates chance level (0.5)

Stage 2: Participants acquired the discrimination over training for both the control cues and the recombined cues, as the accuracy for both recombined compounds and control compounds increased (Figure 31). The two-way ANOVA was applied to test whether the two factors (trial-type and block) had effect on accuracy. It revealed that there was a significant main effect of block [$F(3,60) = 26.69$, $p < 0.001$, $\eta_p^2 = .57$, 90% CI [.41, .65], power = 1.00], but no main effect of trial-type [$F < 1$, $p > 0.99$], and no interaction between factors [$F < 1$, $p = 0.58$].

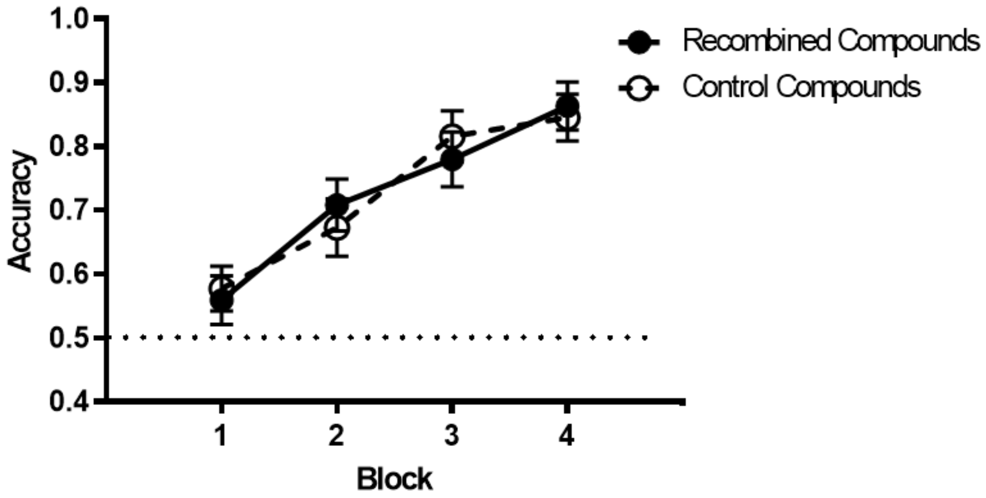

Figure 31. Accuracy in four blocks in stage 2. Error bars indicate ± standard error of mean. And the dash line on the left panel indicates chance level (0.5)

Test phase: The ratings for the test stage are shown in Figure 32. The ratings for compounds consisting of cues paired with outcome 4 (BD and QS) were higher than for those paired with outcome 3 (AC and PR), indicating that participants had learnt the cue-outcome associations. More importantly, the difference between biconditional compounds was greater than the difference between Predictive/Irrelevant compound. A two-way ANOVA (condition: predictive cues (AC, BD) vs. biconditional cues (PR, QS); Outcome: 3 (AC, PR) vs. 4 (BD, QS)) showed that the difference between cues paired with outcomes 3 and 4 was greater for the biconditional condition than the predictive condition. There was a significant main effect of outcome [$F(1,20) = 59.39$, $p < 0.001$, $\eta_p^2 = .75$, 90% CI [.54, .83], power = 1.00], but no main effect of condition [$F < 1$, $p = 0.78$]. There was an interaction between condition and outcome [$F(1,20) = 8.74$, $p = 0.008$, $\eta_p^2 = .30$, 90% CI [.05, .50], power = .84] demonstrating that the effect of outcome was significantly greater for the biconditional cues than for predictive cues. Simple main effect analysis showed that the score of PR compound was significantly lower than the score of AC compound [$F(1,20) = 5.39$, $p = 0.03$, $\eta_p^2 = .21$, 90% CI [.01, .43], power =.64] and the score of QS was significantly higher than the score of BD [$F(1,20) = 6.34$, $p = 0.02$, $\eta_p^2 = .24$, 90% CI [.02, .45], power = .71]. Not surprisingly, the simple effects showed that participants discriminated successfully for both biconditional cues [$F(1,20) = 82.17$, $p < 0.001$, $\eta_p^2 = .80$, 90% CI [.63, .86], power = 1.00] and predictive cues [$F(1,20) = 15.74$, $p = 0.001$, $\eta_p^2 = .44$, 90% CI [.15, .61], power = .98]. There was no difference between compound EH and compound FG [$F(1,20) = 0.82$, $p = 0.42$, $\eta_p^2 = .04$, 90% CI [.00, .23], power = .15], and two one-sample t tests showed that the scores of both compounds (EH and FG) were not significantly different from the score 5 [$t < 1$, $p > 0.4$]. However, there was a significant difference between compound IJ and compound KL [$F(1,20) = 8.69$, $p < 0.001$, $\eta_p^2 = .30$, 90% CI [.05, .50], power = .84].

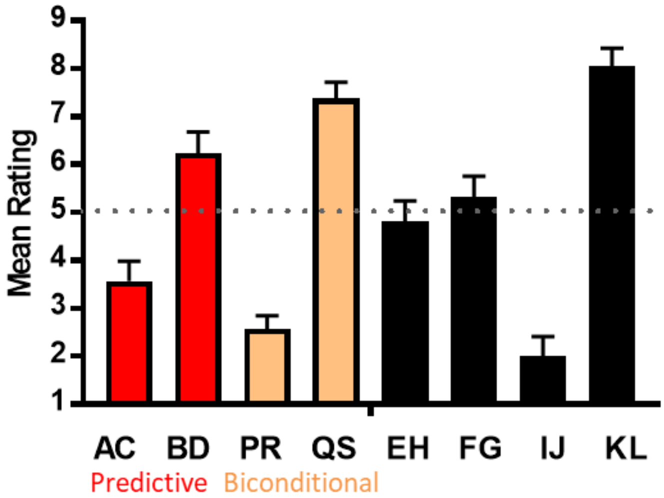

Figure 32. The rating scores of each compound in test phase. The Y axis is the mean rating score. Error bars indicate standard error of mean.

Discussion:

The main finding of the current experiment was that predictive cues received less attention than biconditional cues. Combined with the results of Experiment 9 which showed that biconditional cues receive similar level of attention to uncertain cues, we can conclude that biconditional cues functioned in a similar manner of uncertain cues in the current experiment. Once again, the summed prediction error could not explain current results. The summed prediction errors of biconditional cues and predictive cues were both low, but the results showed attention paid to biconditional cues was higher than predictive cues. Instead, the individual prediction error is suitable to explain this data. The individual prediction error of biconditional cues was greater than the individual prediction error of predictive cues.

Experiment 11

Introduction:

From Experiment 1-8, I have found that task difficulty modulates attention. When the task (Experiment 2 & Experiment 4) was relatively easy (4 uncertain compounds & 8 certain compounds), attention paid to uncertain cues was greater than predictive cues, while predictive cues received more attention when the task (Experiment 5 & Experiment 6) was relatively difficult (8 uncertain compounds & 8 certain compounds). Experiment 9 and Experiment 10 suggested that biconditional cues functioned in a similar manner to uncertain cues. If this argument is true, then it could be anticipated that under the simple training procedure as demonstrated in Experiments 2 and 4, attention paid to biconditional cues should be higher than attention paid to predictive cues. That is exactly what I have found in Experiment 10. The training procedure of Experiment 10 included 8 certain compounds and 4 biconditional compounds, in which the task difficulty is similar to the simple version (Experiment 2 & Experiment 4). Thus, in the current experiment, the task difficulty was adjusted to the difficult version (the task difficulty being similar to that in Experiment 5 and Experiment 6) by adding four extra uncertain compounds. There were 8 certain compounds, 4 biconditional compounds and 4 uncertain compounds in stage 1. It could be anticipated that predictive cues receive more attention than biconditional cues, if the role of biconditional cues is similar to the role of uncertain cues.

Participants:

Twenty-four people (4 males and 20 females) participated in the experiment. The age range was 18-36 (mean: 20.9, SD: 4.1). All participants had normal or corrected to normal vision. Durham University Psychology students received participant pool credit and other participants were compensated for their time at a rate of £10/hour.

Procedures:

In stage 1, there were three types of cue compounds (see Table 11): certain compounds (AV→1, BV→2, AW→1, BW→2), uncertain compounds (ZM→1/2, ZO→1/2, NO→1/2, NM→1/2) and biconditional compounds (PQ→1, PS→2, RS→1, RQ→2). It is similar to Experiment 10 except for four extra uncertain compounds. The procedure in stage 2 and the test phase were the same to the procedure of Experiment 10. Each of the first four compounds of stage 2 were composed of one predictive cue (A,B,C,D) and one biconditional cue (P,Q,R,S) from stage 1. All other details were the same as previous experiments.

Table 11. Design of Experiment 11. Letters represent cues (presented in compound) and numbers represent outcomes. During the test stage, participants were asked to rate the expected likelihood of outcome 3 or outcome 4 given the presented cue compound.

Stage 1	Stage 2	Test
AV → O1	AP → O3	AC → O3/O4?
BV → O2	BQ → O4	BD → O3/O4?
AW → O1	CR → O3	PR → O3/O4?
BW → O2	DS → O4	QS → O3/O4?
CX → O2	EF → O3	EH → O3/O4?
DX → O1	GH→ O4	FG → O3/O4?
CY → O2	IJ → O3	IJ → O3/O4?
DY → O1	KL → O4	KL → O3/O4?
PQ → O1		
PS → O2		
RS → O1		
RQ → O2		
ZM → O1/O2		
ZO → O1/O2		
NO → O1/O2		
NM → O1/O2		

Predictive cues
Irrelevant cues
Biconditional cues
Uncertain cues

Results:

Stage 1: Participants acquired the discrimination over training with performance increasing for the certain compounds and biconditional compounds over blocks, but no improvement for the uncertain compounds (see Figure 33). A two-way ANOVA was carried out to test whether block (1-4) and certainty (biconditional, predictive & uncertain compounds) had an effect on accuracy in stage 1. There was a significant effect of block [$F(3,69) = 7.84$, $p < 0.001$, $\eta_p^2 = .25$, 90% CI [.10, .36], power = .99] and certainty [$F(2,46) = 49.45$, $p < 0.001$, $\eta_p^2 = .68$, 90% CI [.53, .75], power = 1.00], and a significant interaction between block and certainty [$F(6,138) = 6.84$, $p < 0.001$, $\eta_p^2 = .23$, 90% CI [.11, .30], power = 1.00]. Simple main effect analysis showed that there is no difference between certain compound and biconditional compounds ($p > 0.87$), but accuracy of uncertain compounds differed from certain compounds and biconditional compounds (*p*s < 0.001).

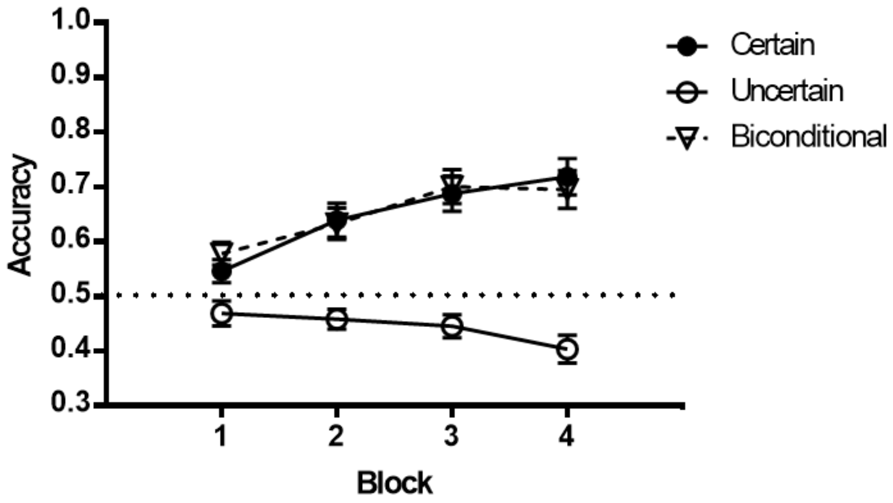

Figure 33. Accuracy across four blocks of stage 1. Error bars indicate ± standard error of mean. The dash line on the left panel indicates chance level (0.5)

Stage 2: Participants acquired the discrimination over training for both the control cues and the recombined cues, as the accuracy for both recombined compounds and control compounds increased (Figure 34). The two-way ANOVA revealed that there was a significant main effect of block on accuracy [$F(3,69) = 30.48$, $p < 0.001$, $\eta_p^2 = .60$, 90% CI [.42, .65], power = 1.00], but no main effect of trial-type [$F(1,23) = 2.47$, $p = 0.13$, $\eta_p^2 = .10$, 90% CI [.00, .30], power = .35] on accuracy. There was no significant interaction between factors [$F < 1$, $p = 0.63$].

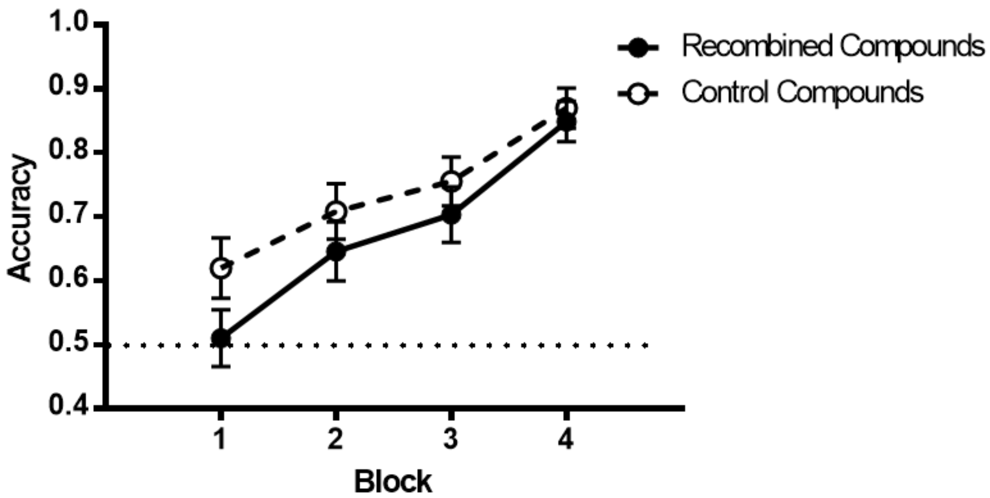

Figure 34. Accuracy in four blocks in stage 2. Error bars indicate ± standard error of mean. The dash line on the left panel indicates chance level (0.5)

Test Phase: The ratings for the test stage are shown in Figure 35. The ratings for compounds consisting of cues paired with outcome 4 (BD and QS) were higher than for those paired with outcome 3 (AC and PR), indicating that participants had learnt the cue-outcome associations. More importantly, the difference between predictive cues (AC,BD) is greater than the difference between biconditional cues (BD, QS). A two-way ANOVA (condition: predictive cues (AC, BD) vs. biconditional cues (PR, QS); Outcome: 3 (AC, PR) vs. 4 (BD, QS)) showed that the difference between cues paired with outcomes 3 and 4 was greater for the predictive condition than the biconditional condition. There was a significant main effect of outcome [$F(1,23) = 60.10$, $p < 0.001$, $\eta_p^2 = .72$, 90% CI [.52, .81], power = 1.00], but no main effect of condition [$F(1,23) = 1.24$, $p = 0.28$, $\eta_p^2 = .05$, 90% CI [.00, .23], power = .20]. There was an interaction between condition and outcome [$F(1,23) = 4.70$, $p = 0.041$, $\eta_p^2 = .17$, 90% CI [.00, .37], power = .58] suggesting that the effect of outcome was significantly greater for the predictive cues than for biconditional cues. Simple main effect analysis showed that the score of the PR compound was significantly higher than the score of the AC compound [$F(1,23) = 6.05$, $p = 0.02$, $\eta_p^2 = .21$, 90% CI [.02, .42], power = .69] but the score of QS was not significantly lower than the score of BD [$F(1,23) = 1.43$, $p = 0.24$, $\eta_p^2 = .06$, 90% CI [.00, .25], power = .22]. Not surprisingly, the simple effects showed successful discrimination for both biconditional cues [$F(1,23) = 15.92$, $p < 0.001$, $\eta_p^2 = .41$, 90% CI [.14, .58], power = .98] and for predictive cues [$F(1,23) = 53.39$, $p < 0.001$, $\eta_p^2 = .70$, 90% CI [.48, .79], power = 1.00]. There was no difference between compound EH and compound FG [$F(1,23) = 0.51$, $p = 0.61$, $\eta_p^2 = .02$, 90% CI [.00, .37], power = .11], and two one-sample t tests showed that the scores of both compounds (EH and FG) were not significantly different from chance [$t < 1$, $p > 0.4$]. However, there was a significant difference between compound IJ and compound KL [$F(1,23) = 11.67$, $p < 0.001$, $\eta_p^2 = .34$, 90% CI [.09, .52], power = .93].

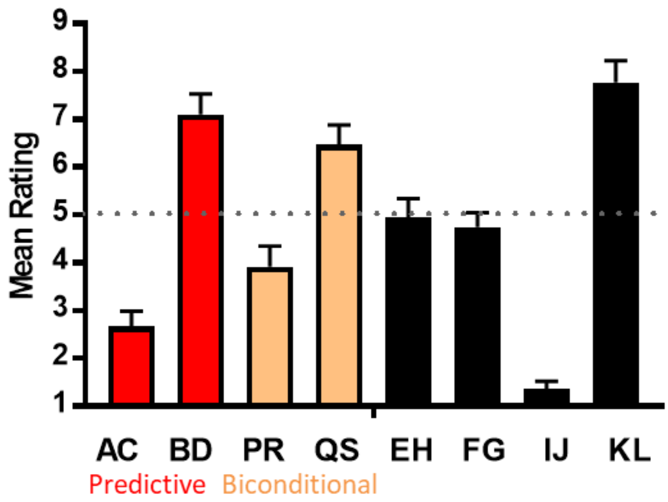

Figure 35. The rating scores of each compound in test phase. The Y axis is the mean rating score. Error bars stand for standard error of mean.

Discussion:

The mean rating scale in the test phase (see Figure 34) showed that the difference between compound AC and BD was greater than the difference between compound PR and QS. These results suggested attention paid to predictive cues was larger than attention paid to biconditional cues. The complexity of training procedure of the current experimental design (8 certain compounds, 4 uncertain compounds and 4 biconditional compounds) was more complicated compared to the task difficulty of Experiment 10 (8 certain compounds and 4 biconditional compounds). Combined with the results of Experiment 10 (biconditional cues received more attention than predictive cues), the complexity of training procedure appears to determine the associability of a cue. In other words, participants paid attention to biconditional cues when the task difficulty was relatively easy, but to predictive cues when the training procedure was relatively difficult. Experiment 11 replicated the main finding of Chapter 3: task difficulty modulates attention, and also verified that biconditional cues functioned in a similar way to uncertain cues.

General Discussion

This chapter was to test whether the associability is modulated by the summed prediction error or the individual prediction error. And whether the manipulation of task difficulty results in different learning effects under the cue comparison between biconditional cues and predictive cues. The results of this chapter showed that the associability of a cue might be driven by the individual prediction error term, and demonstrated that task difficulty determine which type of cues receive more attention when the comparison between biconditional cues and predictive cues was made.

The results of Experiment 9 showed that uncertain cues receive similar attention to biconditional cues, which suggested that the summed error term may not be the crucial factor in determining associability, as the summed error term of uncertain cues is high but the summed error term of biconditional cues is low. Instead, the individual error term may be a vital factor in modulating associability, since the individual error terms of biconditional cues and uncertain cues are both high. That's why they received similar degree of attention. To be further, the comparison between biconditional cues and predictive cues was made in Experiment 10. The summed prediction errors of biconditional cues and predictive cues were both low, but the individual prediction error of biconditional cues (high) is not close to the individual prediction error of predictive cues (low). Biconditional cues received more attention than predictive cues this time, suggesting that the individual error term determine the associability of a cue. Taken together, Experiment 9 and Experiment 10 showed that attention is influenced by the individual error term rather than the summed error term.

Another key finding is that the difficulty of training procedure can determine which type of cues participants pay attention to demonstrated by Experiment 10 and Experiment 11. When the task was relatively easy, participants paid attention to biconditional cues over predictive cues (Experiment 10), while, under the relatively complicated training procedure, attention paid to predictive cues was higher than attention paid to biconditional cues (Experiment 11). The results replicated the main finding of Chapter 3. The task difficulty can modulate attention not only in the comparison between predictive cues and uncertain cues but also in the comparison between predictive cues and biconditional cues, which may suggest that task difficulty play an important role in associative learning.

Livesey et al. (2011) compared biconditional cues to predictive cues under the high task difficulty (8 biconditional compounds, 8 certain compounds), and the results showed that attention paid to predictive cues is higher than biconditional cues, which is similar to my results of Experiment 11 (8 certain compounds, 4 biconditional compounds and 4 uncertain compounds). They did not manipulate the task difficulty, so I could not compare my results to their results when the task difficulty was low (the number of uncertain compounds is less). Combined with the finding of Experiment 5 and Experiment 6 (predictive cues received more attention than uncertain cues under the high task difficulty), I have replicated the results of Livesey et al. (2011). When the task difficulty is high, the learned predictiveness effects were observed no matter what the comparisons were. In Experiment 5 and 6 (the comparison between predictive cues and uncertain cues.), the results showed that predictive cues receive more attention than uncertain cues; In experiment 11 (the comparison between biconditional cues and predictive cues.), attention paid to predictive cues was higher than biconditional cues.

The possible reasons for the task difficulty drive different learning effects will be discussed in the next chapter. Moreover, across all my experiments, the results suggested that learning is modulated by the individual error term. This finding is also support by many other studies (Uengoer,Dwyer, Koenig & Pearce, 2019; Uengoer, Lachnit & Pearce, 2019), and the details will also be discussed in the next chapter.

Chapter 5

5.1 General Discussion

Depending on the nature of the training procedure, both predictiveness effects (Chapter 3) and uncertainty effects (Chapter 2) were observed. Importantly, the crucial factor that determined which of these effects was present in a particular experiment was found. By changing the task difficulty, it was possible to switch the resultant learning effect. When the task difficulty was relatively low, attention was preferentially allocated to uncertain cues, but more attention was paid to predictive cues when the task difficulty was relatively high. Chapter 4 showed that biconditional cues functioned in a similar way to uncertain cues, in that biconditional cues received more attention than predictive cues and received similar levels of attention to uncertain cues. Moreover, the task difficulty of the training procedure also altered the allocation of attention between biconditional cues and predictive cues. When the complexity of the training procedure was high, more attention was allocated to the predictive cues, but when the complexity of the training procedure was low, attention was shifted to the biconditional cues. Based on a series of experiments, it is concluded that learned predictiveness and learned uncertainty might be governed by individual prediction error, rather than summed error. This finding will be returned to discuss in the later part.

Chapter 3 demonstrated that the learned predictiveness and learned uncertainty might be determined by the task difficulty of training procedure. When the training procedure was difficult, attention was paid to predictive cues; while attention was allocated to uncertain cues when the training procedure was relatively easy. There are few possible explanations for this finding. Firstly, task difficulty may drive the application of different attentional strategies. In other words, different attentional strategies (attentional exploitation & attentional exploitation) were applied to the relatively simple training procedure (Experiment 2 and 4, 4 uncertain compounds) and the relatively complex training procedure (Experiment 5 and 6, 8 uncertain compounds). Perhaps, the task difficulty influenced the processes of exploitation and exploration. The dilemma of exploitation-exploration in the process of decision-making has been proposed for decades (March, 1991; March, 2006). This trade-off could be determined by some factors such as novelty of task (Krebs, Schott, Schütze & Düzel, 2009), the benefit of using exploitation and the cost of applying exploration (Cohen, McClure, & Yu, 2007). If the benefit of using exploitation is bigger than the cost of using exploration, then exploitation outweighs exploration, and vice versa. The process of association between cues and outcomes can be considered as decision-making, as participants have to learn the relationship between cues and outcomes, and decide which cue they need to pay attention to. Beesley et al., (2015) pointed out two attentional strategies: attentional exploration and attentional exploitation might coexist within an experiment. Attentional exploration is a strategy in which participants allocate their attention to the cues that unreliably link to outcomes in order to reduce uncertainty; while attentional exploitation suggests that attention is allocated to the most informative cues. In the presented experiments of this thesis, which different attentional strategies are applied depends on the different task difficulties. When the complexity of training procedure was relatively high, participants had to allocate tremendous cognitive resources to explore uncertain cues, which might not be an efficient way to finish the task. Thus, exploitation strategy is better than exploration strategy given situation of high difficulty of task. In contrast, when the complexity of training procedure was relatively low, participants have extra cognitive resources to reduce uncertainty by paying attention to uncertain cues. In this way, exploration might be a better strategy. In this way, task difficulty might influence the trade-off between exploitation and exploration during the learning process of cue-outcome association.

Secondly, another possible explanation to the findings is that task difficulty can determine the two types of CS processing: controlled attention and automatic attention. Automatic attention is a relatively fast process, involving less participant effort. However, controlled attention is a relatively slow process, involving more participant effort and providing a large degree of participant control (Schneider, Dumais, & Shiffrin, 1982). The processing of controlled attention will be applied when the presented stimulus is novel, or when the relationship between stimuli and outcomes is uncertain. This controlled attention increases when the relationship between cues and subsequent outcomes is unreliable. However, when participants/animals are familiar with the task or the relationship between cues and outcomes is predictive, automatic attention will be applied. The degree of task difficulty can determine the learned predictiveness and learned uncertainty might due to the application of controlled attention. When the complexity of training procedure increases, the effect of controlled attention may decrease. While controlled attention increases when the degree of task difficulty is low. It could be the reason participants applied different forms of attention (automatic attention & controlled attention) to different training procedures based on the task difficulty. Therefore, participants applied automatic attention when a task is well learnt. However, a controlled CS processing will be engaged when the task is still not well learnt. In order words, when the training procedure was relatively easy (Experiment 2 and 4), automatic attention was applied; while controlled attention was involved when the training procedure was relatively difficult (Experiment 5 and 6). Applying controlled attention or automatic attention might be driven by the degree of task difficulty.

Finally, the other possibility to explain the results is that unexpected uncertainty can draw more attention than expected uncertainty. Easdale, Le Pelley and Beesley (2017) found the onset of uncertainty can boost the learning rate of a new cue-outcome relationship. In their study, there were two groups in terms of the level of uncertainty in stage 1: certain group and uncertain group. For the certain group, cue compounds consistently led to specific outcomes (AX→ 1 (100%), AY → 1 (100%), BX → 2 (100%), BY → 2 (100%)); while, for the uncertain group, cue-outcome relationship was probabilistic (AX→1 (80%), AX →2(20%); AY→1(80%), AY→2(20%); BX→1(20%), BX→2(80%); BY→1(20%), BY →2(80%)). In stage 2, all participants experienced a new cue-outcome relationship, which was also probabilistic. The results showed that in terms of stage 2 learning, participants in the certain group from stage 1 learned faster than participants in the uncertain group. Due to the stage 1 training, participants from certain group experienced an unexpected onset of uncertainty in stage 2, while participants from the uncertain group consistently experienced the same degree of uncertainty. Therefore, participants from the certain group allocated more attention than participants from the uncertain group to uncertain cues in stage 2. The key point of this study was that unexpected uncertainty facilitates learning compared to expected uncertainty. Easdale et al., (2017) suggested that the learning rate of a cue is determined by the experiences of uncertainty. In other words, the uncertain effect is dynamic. When uncertainty becomes expected uncertainty, attention paid to the uncertainty will be low; while attention paid to uncertain cues will be high when the uncertainty is still unexpected. Perhaps this idea can explain different learning effects can be driven by task difficulty. In the presented experiments, when the training procedures were complicated (8 uncertain compounds), participants experienced a more uncertain cue-outcome relationship. Therefore, unexpected uncertainty became expected uncertainty. Consequently, attention paid to uncertain cues was low. In contrast, when the training procedures were simple (4 uncertain compounds), the uncertain cue-outcome relationship still remained unexpected. Therefore, attention paid to unexpected uncertain cues was high. In a nutshell, the unexpected uncertainty and expected uncertainty might be governed by task difficulty.

From the series of experiments (Experiment 2- Experiment 8), task difficulty may play an important role in learned predictiveness and learned uncertainty. The possible mechanisms were discussed above. The concept of task difficulty has been discussed in previous literatures (Lavie, Hirst, De Fockert & Viding, 2004; Lavie, 2005; 2006; 2010), in which task difficulty can influence cognitive load (cognitive control). In those studies, the degree of task difficulty was defined by the levels of working memory. Lavie et al., (2004) proposed a load theory of attention to describe distractor rejection may rely on the different types of load (perceptual load and cognitive load) and levels of load (high and low). According to their studies, high level of perceptual load diminished the interference of distractors, whereas low level of cognitive load enhanced the interference of distractors. For the perceptual load, the most commonly used paradigm is the response competition procedure (Forster &

Lavie, 2007) and the attentional capture procedure (Forster & Lavie, 2008). For example, Forster and Lavie (2007) used the response competition procedure to test whether distractors can interfere participants' attention under the low and high perceptual load. The visual stimulus was consisted of six letters which were presented as a circle with 1.6° radius. Participants were requested to search the target letter (letter N or X) from this circle. For the high perceptual load condition, the other five non-target letters were randomly selected from letters (H, K M, V, W, Z); while the other five non-target letters were letter O in the low perceptual load condition. In other words, under the low perceptual load, the target letter (either N or X) appeared as the pop-out effect; but participants had to put extra effort to search the target letter under the high perceptual load. A distractor letter was presented to either left side of circle or right side of circle. There were two types of distractors: congruent distractors and incongruent distractors. The congruent distractor was the distractor was the same as the target letter; while the incongruent distractor letter differed to the target letter. The result showed that the difference between reaction time of congruent condition and no distractor condition was longer than the difference between reaction time of incongruent condition and no distractor condition when the perceptual load was low. This difference was diminished when the perceptual load was high. It suggested that distractor interference increased under the low perceptual load but decreased when the perceptual load was high.

For the cognitive load, Lavie et al., (2004) manipulated the level of working memory to test whether distractors interfere the attentional task. In this task, participants needed to memorize a set of digits (high level of load: six digits; low level of load:

one digit) in the beginning of each trial. Then participants were instructed to make manual response based on a target letter (either x or z) and ignored the distractors (capital X, Z or N). The location of the target letter was either on the left side or right side of the centre of the screen; while the location of the distractor was either above or below the centre of the screen. There were three conditions: compatible (target was x, distractor was capital X, or target was z, distractor was capital Z), incompatible (target was x, distractor was capital Z, or target was z, distractor was capital X) and neutral condition (distractor was N no matter what the target was). By the end of trial, a probe digit was presented, and participants were requested to specify whether the digit was presented or absent. The results showed that the difference between reaction time of incompatible condition and neutral condition was longer than the difference between reaction time of compatible condition and neutral condition under the high cognitive load. When the cognitive load was low, this difference was significantly decreased. In other words, the distractors interfered the task performance when the cognitive load was high. Similar results were found in other empirical studies (e.g., Lavie & De Fockert, 2005). Lavie's findings suggested that there were two attentional mechanisms: one is that the perceptual attention reduces the interference effect from distractors when the perceptual load is high, as high perceptual load exhausts the capacity of dealing relevant information or stimuli; The other is the cognitive control mechanism. Cognitive control allows that the attention paid to relevant stimuli overweight the attention paid to irrelevant stimuli (even the irrelevant stimuli are already perceived) when the load of cognition is low; while cognitive control could not maintain high level attention to deal with relevant stimuli and ignore the distractor interference under the high cognitive load.

From the point of Lavie's load theory, task difficulty drove different learning effects might be explained. The visual stimulus of each trial in training stage was composed of two flags and two outcomes for both easy group (4 uncertain compounds) and difficult group (8 uncertain compounds). In other words, the perceptual load in both groups should be the same. However, the cognitive load of difficult group should be higher than the load of easy group due to the difference of the number of uncertain compounds. The uncertain cues could possibly interfere the association between cues and outcomes, as uncertain cues were partially reinforced. When the cognitive load (task difficulty) was low, participants could still maintain high level cognitive control (working memory) to deal with uncertain information. Thus, attention paid to uncertain cues was high; while participants could not maintain high level of cognitive control to deal with the interference of uncertain information under the high cognitive load (difficult group). Therefore, attention paid to predictive cues was higher than uncertain cues.

So far, the reasons for task difficulty driving the different learning effects have been discussed. A further question concerns the exact nature of what constitutes task difficulty. The difference between the simple training procedure (four uncertain compounds, eight certain compounds) and the complex training procedure (eight uncertain compounds, eight certain compounds) is the number of uncertain compounds. The difference in the observed learning effects could have resulted from the increased cognitive memory load or the increase in the number of uncertain compounds. Experiment 8 was designed to answer this question. The stage 1 training procedure in Experiment 8 contained twelve certain compounds and four uncertain compounds. If the task difficulty was only a result of the number of uncertain compounds, then the results should have shown that uncertain cues received more attention than predictive cues (as in Experiment 2 and 4), because the number of uncertain compounds in Experiment 8 was the same as the number of uncertain compounds in Experiments 2 and 4. However, the results showed that predictive cues received a similar level of attention to uncertain cues. Experiment 8 failed to show that uncertain cues received more attention than predictive cues, and also failed to show the opposite finding. The task difficulty may therefore be dependent on both the number of uncertain compounds and the number of cues. In terms of the level of task difficulty, Experiments 5 and 6 were the complex version (eight certain compounds and eight uncertain compounds) and Experiments 2 and 4 were the simple version (eight certain compounds and four uncertain compounds). The difficulty of the training procedure in Experiment 8 (twelve certain compounds and four uncertain compounds) was lower than the difficulty of the complex version, but it was higher than the task difficulty of the simple version. This idea, however, needs more empirical evidence which includes a systematic manipulation of task difficulty to further investigate the relationship between cues and outcomes.

Although I consistently found the complexity of experimental design can drive different learning effects (learned predictiveness and learned uncertainty), there were still some studies describing that the learned predictiveness effects were obtained under simple training procedures, which were not consistent with my findings: the learned predictiveness effects were obtained under the complex training procedures. Before I discuss counterexamples, the supportive studies are discussed. Hall and Pearce (1979) found the negative transfer effect in animal study, in which the cue paired with the minor reaction in the first stage should weaken the associative strength of the cue paired with the strong reaction in the second stage. Griffiths et al., (2011) used an allergist task to find the uncertainty effect in which the uncertain relationship between cues and outcomes increases attention by using similar procedure in human study. In the first stage, the cue was paired with an outcome (with the minor allergic reaction: +). In the second stage, the same cue now was paired with another outcome (a strong allergic reaction: ++). They found the negative transfer effect. However, this effect was diminished by a surprising event occurs between stage 1 and stage 2. In Griffiths' following experiment, half participants were presented surprising trials between two stages (the change group), in which the cues were linked to no outcome. The other participants were presented the same procedure as the design of the previous experiment (the negative transfer group). The results showed that the cue of the change group received more attention than the cue of the negative transfer group in stage 2, which supports the uncertainty principle (a release form negative transfer). The task difficulty of Griffiths et al (2011) was low, as there are only three cues in the training stage. It could be an evidence for my findings. However, it can not be fully replicated by many other studies (e.g., Packer, Siddle & Tipp, 1989). They used similar procedures, but they failed to find the uncertainty effect.

Another example is that Hogarth et al., (2008) used a simple training procedure: AX+ CX- and BX+/- (only two certain compounds: AX+, CX- and one uncertain compound: BX+/-), and the uncertainty effect was observed. It seems that the study is consistent to my findings. Austin and Duka (2010, 2012), however, tried to replicate the study of Hogarth et al (2008), but they failed to reproduce the similar results. In Hogarth's designs, there were three types of trials: the first trial type was that a cue compound composed of two visual cues (A and X) always led to an auditory outcome (AX+), the second was that a visual cue compound (B and X) equally led to an auditory outcome or no outcome (BX +/-), and the other was that a visual cue compound (C and X) reliably led to no outcome (CX-). Hogarth found the pupils dwelling time on each cue was B > A = C, which suggested that the overt attention on uncertain cues (B) was larger than predictive cues (A). However, Austin and Duka (2010) used a similar paradigm (different sound for outcome), and the totally opposite results (A > B > C) were obtained. Later on, Austin and Duka (2012) again found the pupil dwelling time on predictive cues (A) was longer than uncertain cues (B). In this case, the experimental design was extremely simple (only three trial types), which is even easier than my simple version (Experiment 2 and 4, 12 trial types). Taken together, all studies (Hogarth et al., 2008; Griffiths et al., 2011 & the presented experiments of this thesis) which included the uncertainty effect are simple training procedures, although Hogarth's study and Griffiths' study are not fully replicated by other studies.

Compared with the studies of Austin and Duka (2010, 2012), there were some procedure differences between their studies and my experiments. For example, Austin and Duka used visual stimuli as cue compounds and audio stimuli as outcomes. Only visual stimuli were utilized as cues and outcomes in my experiments. It should be noted that there is no literature directly describing the difference of stimuli causes different learnings. Thus, it is still unknow whether this factor results in different learnings. The other difference is that overt attention was measured as an index for learning in their studies. Instead, learning in a subsequent training stage to assess the effect of prior predictiveness or uncertainty was used in my experiment. It is very difficult to make comparisons across two studies given these fundamental differences. With regarding to the differences, there were studies which overcame these differences still showing that the learned predictiveness effects under the simple training procedures. For instance, Le Pelley et al., (2010) used a simple procedure which includes predictive cues and uncertain cues, and found the learned predictiveness effect. In stage 1, six trial types (A → pink, B → orange, C → orange, D → pink, X → pink/ orange, Y → pink/orange) were presented. A,B,C,D were predictive cues, and X,Y were uncertain cues. In stage 2, each cue (A,B,C,D,X,Y) reliably link to a specific new outcome (A+, B-, C+, D-, X+, Y-). In the test phase, participants needed to choose one cue base on nine different comparisons (e.g., A vs. B, C vs. D……). The results showed that predictive cues received more attention than uncertain cues, which was contradict to my experiments. Moreover, Le Pelley et al., (2010) only used visual stimuli as cues and outcomes, and used the learning in a subsequent training stage as an index to examine the effect of prior predictiveness or uncertainty, which is similar to my experimental design but different from Austin and Duka (2010, 2012) studies. Once again, they found the learned predictiveness effect. In this case, the task difficulty might not be a critical factor for determining the effects of learned predictiveness and learned uncertainty, as the task difficulty (two uncertain cues) of the study of Le Pelley et al. was even lower than the simple version of my experiments (4 uncertain compounds). Nevertheless, there were still few differences between Le Pelley's designs and my experimental procedures. Firstly, Le Pelley used a single cue to pair with an outcome (e.g., A → pink), but cue compounds were utilized in my experiment (e.g., AX → 1). Secondly, the test phase in Le Pelley's study asked participants to make a choice based on a comparison (e.g., A vs. B), but the test phase in my experiments asked participants to make a prediction based on the presented cue compound (e.g., how likely the cue compound (AC) led to outcome 3 or outcome 4). Another possibility to account the divergent results between Le Pelley's studies and my studies is that the relationship between task difficulty and the learned uncertainty might not be linear. For the presented thesis experiments, the results indicated that learned predictiveness effects were obtained when the task is relatively easy; while the learned uncertainty effects were obtained when the task is relatively complicated. If the relationship between the learning effects and the task difficulty is non-linear, it could possible explain other studies (e.g., Le Pelley (2010), Austin and Duka (2010, 2012)): when the task is easy, attention paid to predictive cues is high. Form this point of view, when the task difficulty is really easy (e.g., only a single cue), attention is allocated to predictive cues (e.g., Le Pelley et a., 2010); while, when the task difficulty is high (e.g., Livesey et al., 2011; Experiment 5 and Experiment 5 of this thesis, eight uncertain compounds), attention paid to predictive cues is also high. However, when the task difficulty is in between, attention paid to uncertain cues is higher than predictive cues. The possible explanation is that at the specific level of task complexity, attention paid to uncertain cues will be high. So far, none of the current attentional models would predict such a non-linear relationship. The discrepancy between the present experiments and other studies needed to be further investigated.

It should be noted that although the studies (Griffiths et al., 2011; Hogarth et al., 2008) that report the uncertainty effect have not been fully replicated, the task difficulty of their training procedures were low (three cues in Griffiths et al.' study and three compounds in Hogarth et al.'s). In these cases, the level of difficulty of the training procedure might be a crucial factor in determining the learning effect that is observed. The low difficulty of the training procedure might encourage participants to pay more attention to the uncertain cues. However, if the task difficulty increases, participants might give up paying attention to uncertain cues in favour of paying attention to the predictive cues. Watson, Pearson, Chow, Theeuwes, Wiers, Most, and Le Pelley (2019) combined a visual search task with different levels of memory loading (task difficulty) to investigate the role of cognitive control (controlled process) in attentional capture. In the visual search task, participants were instructed to ignore the colour distractor (blue or orange) and make a saccadic eye-movement to the target location (a diamond shape). Before the visual search task, they manipulated the task difficulty by altering the memory loading. Either five digits were presented (high task difficulty) or one digit was presented (low task difficulty). After participants had made a response to the target location, they then needed to recall the digit(s) that preceded the visual search task. The results showed that participants' attention was captured by a distractor under the high degree of task difficulty more so than under the low degree of task difficulty. It was suggested that the cognitive control (controlled attention) was sabotaged under the high memory loading, which may be in support of the findings in this thesis. Thus, the controlled attention paid to uncertain cues decreased when the task difficulty was high, whereas the controlled attention increased when the task was relatively easy.

Based on the Pearce-Hall model, the summed prediction error can determine how much attention is paid to a cue. In Experiment 2 (predictive cues vs. uncertain cues) and Experiment 3 (irrelevant cues vs. uncertain cues), the results showed that uncertain cues received more attention than irrelevant cues and predictive cues, which might be explained by the Pearce-Hall model. The summed prediction error of uncertain cues was greater than the summed prediction error of irrelevant cues and predictive cues. It seems that attention was controlled by the total prediction error for the entire compound rather than the individual prediction error for each cue, which would have been the prediction of the Mackintosh model. However, for Experiment 3, there are a few possible reasons why the uncertain cues received more attention than the predictive cues. Firstly, the uncertain compound was composed of two uncertain cues, while the certain compound was composed of one predictive cue and one irrelevant cue. As the predictive cue perfectly predicted the outcome, participants could focus on the predictive cues to solve the discriminations. Le Pelley et al. (2011) also suggested that overt attention paid to predictive cues was greater than to irrelevant cues. Therefore, when the comparison was made between uncertain cues and irrelevant cues, attention paid to the uncertain cues was greater than attention paid to the irrelevant cues. In other words, the uncertainty effect was due to the learning of the predictive cues. Secondly, participants might have learned not to look at the irrelevant cues. Since the irrelevant cues were redundant, the best way to deal with those cues was to ignore them. Therefore, the uncertainty effect might have been due to the irrelevant cues being ignored. As such, whether the summed prediction error or the individual prediction error was responsible for driving the associability needed to be further investigated. Chapter 4 (Experiments 9 - 11) provides an answer this question. Experiment 9 compared biconditional cues to uncertain cues, and the results revealed that the attention paid to biconditional cues was similar to the attention paid to uncertain cues. However, the summed prediction error of biconditional cues is different to the summed prediction error of uncertain cues, therefore summed prediction error might not be able to explain these data. Moreover, Experiment 10 examined the comparison between predictive cues (for which summed prediction error is low) and biconditional cues (for which summed prediction error is also low), and the results showed that biconditional cues received more attention than predictive cues. Once again, summed prediction error can not explain these data. Both of these experiments suggest that the summed error term is not the crucial factor that affects the associability of cues. On the contrary, individual prediction error seems to explain these effects. In Experiment 9, the individual prediction errors for both biconditional cues and uncertain cues are high, as individually they are partially reinforced. Therefore, biconditional cues should have received a similar level of attention to uncertain cues, which was found to be the case. Similarly, in Experiment 10, the individual prediction error of biconditional cues was greater than the individual prediction error of predictive cues, and the results showed that biconditional cues received more attention than predictive cues. These data suggest that the individual prediction error rather than the summed prediction error determined how much attention participants allocated to a cue. Individual prediction error can also explain the observed uncertainty effect in Experiment 2. Thus, the uncertain cues received more attention than the predictive cues as the individual prediction error of uncertain cues is greater than predictive cues.

Based on my experiments, the individual prediction error can affect how much attention participants pay to a cue. There are also some studies describing the individual prediction error can drive the learning effects (Uengoer, Lotz & Pearce, 2013; Uengoer et al., 2019). For instance, Uengoer, Lachnit & Pearce (2019) examined the role of redundant cues in human associative learning. In their task, four trial types (3AX+, BX−, CY+, 3DY−; + indicates that the outcome occurs, -

indicates the absence of the outcome) were presented to participants in stage 1. During the test stage, cues (A,B,C,D,X,Y) were presented to test how likely the cue predicted the outcome. Cues A-D were informative cues, from which participants could predict the outcome. However, cues X and Y were redundant (uninformative) cues. If the associability was driven by the summed prediction error, then the rating of cue X should be similar to the rating of Y. That is because the associative change of X in stage 1 is similar to the associative change of Y. In contrast, if the learning effect is governed by the individual prediction error, then cue Y should be considered a weaker predictor of the outcome than X. The results showed that the associability of cue X was bigger than Y, which is consistent with the theory of individual prediction error.

However, one of the experiments presented in this thesis (Experiment 3) was not consistent with the individual prediction error determining the learning effect. Experiment 3 compared uncertain cues to irrelevant cues. The individual prediction errors of uncertain cues and irrelevant cues were high, but uncertain cues received more attention than irrelevant cues. So far, none of the established learning theories are able to explain all of the learning effects observed. For instance, individual prediction error cannot explain the overexpectation effect (Kamin & Gaioni, 1974), and summed prediction error cannot explain the redundancy effect (e.g., Uengoer et al., 2013). This thesis has provided a lot of evidence for the theory of individual prediction error, although it is still not possible to explain the results of all of the presented experiments using this theory. Which learning theory (summed prediction error or individual prediction error) governs learning effects is still an ongoing topic.

It is worth to noticing that the learning effects in this thesis may relate to inter-trial priming effect (Maljkovic & Nakayama, 1994; Thomson & Milliken, 2013). Maljkovic & Nakayama (1994) found the priming of pop-out effect in which the searching time of singleton is faster when the visual stimuli (the target and distractors) are repeatedly presented in successive trials than when they are not presented in successive trials. In the task, participants were required to find the colour singleton (e.g. red diamond) from the presented stimuli (e.g. red diamond and green diamonds). The results showed that the performance was better when the visual stimuli (the target and distractors) of current trial was the same as previous trial than when they were not. This inter-trial priming may influence the associability as well. For instance, under the reliable association between cues and outcomes (e.g. the combination of U.S. flag and U.K. flag consistently lead to apple outcome rather than bomb outcome), participants chose the apple and got positive feedback. Then, in the following same trial (U.S. flag and U.K. flag were still presented), participants would choose the apple as an outcome, given the previous trial priming. In my thesis, all the trial sequences were randomized. Thus, the inter-trial priming effect should be diminished.

In conclusion, the presented experiments provide robust evidence for not only the learned predictiveness effect but also the learned uncertainty effect. Crucially, individual prediction error seems to be the primary mechanism to determine how much attention participants pay to a cue. Given it is difficult to find the uncertainty effect in human associative learning task, this study may provide some valuable information. An essential goal of future research might be to investigate the neural mechanism of learned predictiveness and learned uncertainty. Constructing the brain circuit for different learning effects provides significant insights into cognitive abilities in humans. Importantly, studying the mechanism of associative learning has implications for the understanding of impaired cognitive function in neuropsychiatric and neurodegenerative disorders. Moreover, the factor of task difficulty was identified to be the key to determining which learning effect was observed. Future research may further focus on what exactly the task difficulty is. For example, whether the time pressure can influence learning. Does limited time increase the difficulty of task? If so, it could possibly be anticipated that attention paid to predictive cues is high under the time pressure. The direction of studying task difficulty may also provide an important insight of learning and attention in a real-world setting, as it is full of all kinds of information whether it is reliable or unpredictable in our daily life.

References

Austin, A. J., & Duka, T. (2010). Mechanisms of attention for appetitive and aversive outcomes in Pavlovian conditioning. *Behavioural brain research*, *213*(1), 19-26.

Austin, A. J., & Duka, T. (2012). Mechanisms of attention to conditioned stimuli predictive of a cigarette outcome. *Behavioural brain research*, *232*(1), 183-189.

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

Bennett, C. H., Wills, S. J., Oakeshott, S. M., & Mackintosh, N. J. (2000). Is the context specificity of latent inhibition a sufficient explanation of learned irrelevance?. *The Quarterly Journal of Experimental Psychology: Section B*, *53*(3), 239-253.

Boll, S., Gamer, M., Gluth, S., Finsterbusch, J., & Büchel, C. (2013). Separate amygdala subregions signal surprise and predictiveness during associative fear learning in humans. *European Journal of Neuroscience*, *37*(5), 758-767.

Broadbent, D. E. (1958). Perception and communication. Elmsford, NY, US.

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological review*, *58*(5), 313.

Cheal, M., & Lyon, D. R. (1991). Central and peripheral precuing of forced-choice discrimination. *The Quarterly Journal of Experimental Psychology*, *43*(4), 859-880.

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933-942.

Cole, G. G., & Liversedge, S. P. (2006). Change blindness and the primacy of object appearance. *Psychonomic Bulletin & Review*, *13*(4), 588-593.

Cole, G. G., & Kuhn, G. (2009). Appearance matters: Attentional orienting by new objects in the precueing paradigm. *Visual Cognition*, *17*(5), 755-776.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, *3*(3), 201-215.

Dickinson, A., Hall, G., & Mackintosh, N. J. (1976). Surprise and the attenuation of blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*(4), 313.

Duffaud, A. M., Killcross, S., & George, D. N. (2007). Optional-shift behaviour in rats: a novel procedure for assessing attentional processes in discrimination learning. *The Quarterly Journal of Experimental Psychology*, *60*(4), 534-542.

Easdale, L. C., Le Pelley, M. E., & Beesley, T. (2017). The onset of uncertainty facilitates the learning of new associations by increasing attention to cues. *The Quarterly Journal of Experimental Psychology*, (just-accepted), 1-49.

Estes, W. K. (1969). Reinforcement in human learning. *Reinforcement and behavior*, 63-94.

Estes, W. K. (1984). Human Learning and Memory. *The Biology of Learning*, 617–628

Eriksen, C. W., & James, J. D. S. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & psychophysics*, *40*(4), 225-240.

George, D. N., Duffaud, A. M., & Killcross, S. (2010). Neural correlates of attentional set. *Attention and associative learning: from brain to behaviour (eds Mitchell CJ, Le Pelley ME)*, 351-383.

Gibson, E. J. (1969). Principles of perceptual learning and development.

Griffiths, O., Johnson, A. M., & Mitchell, C. J. (2011). Negative transfer in human associative learning. *Psychological science*, *22*(9), 1198-1204.

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.

Hall, G., & Pearce, J. M. (1982). Changes in stimulus associability during conditioning: Implications for theories of acquisition. *Quantitative analyses of behavior*, *3*, 221-239.

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.

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

Itti, L., & Koch, C. (2001). Feature combination strategies for saliency-based visual attention systems. *Journal of Electronic imaging*, *10*(1), 161-169.

James, W. (1890). The principles of *Psychology*, *2*, 94.

Kamin, L. J. (1968). Attention-like processes in classical conditioning.

Kamin, L. J., & Gaioni, S. J. (1974). Compound conditioned emotional response conditioning with differentially salient elements in rats. *Journal of Comparative and Physiological Psychology, 87*(3), 591–597.

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.

Krebs, R. M., Schott, B. H., Schütze, H., & Düzel, E. (2009). The novelty exploration bonus and its attentional modulation. *Neuropsychologia*, *47*(11), 2272-2281.

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.

Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*(3), 339.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in cognitive sciences*, *9*(2), 75-82.

Lavie, N., & De Fockert, J. (2005). The role of working memory in attentional capture. *Psychonomic bulletin & review*, *12*(4), 669-674.

Lavie, N. (2006). The role of perceptual load in visual awareness. *Brain research*, *1080*(1), 91-100.

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current directions in psychological science*, *19*(3), 143-148.

Lawrence, D. H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology*, *39*(6), 770.

Lawrence, D. H. (1950). Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *Journal of Experimental Psychology*, *40*(2), 175.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology Section B*, *57*(3b), 193-243.

Le Pelley, M. E. (2010). The hybrid modeling approach to conditioning. *Computational models of conditioning*, 71-107.
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.
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., & Johnson, A. M. (2013). Outcome value influences attentional biases in human associative learning: Dissociable effects of training and instruction. *Journal of Experimental Psychology: Animal Behavior Processes*, *39*(1), 39.
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.
Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature neuroscience*, *14*(10), 1250.
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.
Lubow, R. E., & Moore, A. U. (1959). Latent inhibition: the effect of nonreinforced pre-exposure to the conditional stimulus. *Journal of comparative and physiological psychology*, *52*(4), 415.
Lashley, K. S. (1929). Brain mechanisms and intelligence: A quantitative study of injuries to the brain. University of Chicago Press.
Mackintosh, N. J., & Turner, C. (1971). Blocking as a function of novelty of CS and predictability of UCS. *The Quarterly Journal of Experimental Psychology*, *23*(4), 359-366.
Mackintosh, N. J. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. *Psychological review*, *82*(4), 276.
Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & cognition*, *22*(6), 657-672.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization science*, *2*(1), 71-87.

March, J. G. (2006). Rationality, foolishness, and adaptive intelligence. *Strategic management journal*, *27*(3), 201-214.

Mitchell, C. J., Griffiths, O., Seetoo, J., & Lovibond, P. F. (2012). Attentional mechanisms in learned predictiveness. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*(2), 191.

Pavlov, I. P. (1927). Conditioned reflexes, translated by GV Anrep. *London: Oxford*.

Packer, J. S., Siddle, D. A., & Tipp, C. (1989). Restoration of stimulus associability, electrodermal activity, and processing resource allocation. *Biological psychology*, *28*(2), 105-121.

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.

Pearce, J. M., Dopson, J. C., Haselgrove, M., & Esber, G. R. (2012). The fate of redundant cues during blocking and a simple discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*(2), 167.

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.

Ploghaus, A., Tracey, I., Clare, S., Gati, J. S., Rawlins, J. N. P., & Matthews, P. M. (2000). Learning about pain: the neural substrate of the prediction error for aversive events. *Proceedings of the National Academy of Sciences*, *97*(16), 9281-9286.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of experimental psychology: General*, *109*(2), 160.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive psychology*, *51*(1), 1-41.

Rescorla, R. A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin*, *72*(2), 77.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, *2*, 64-99.

Rescorla, R. A. (2000). Associative changes in excitors and inhibitors differ when they are conditioned in compound. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*(4), 428.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological review*, *84*(1), 1.

Schneider, W., Dumais, S. T., & Shiffrin, R. M. (1982). *Automatic/Control Processing and Attention* (No. HARL-ONR-8104).

Shanks, D. R., & Darby, R. J. (1998). Feature-and rule-based generalization in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*(4), 405.

Shepp, B. E., & Schrier, A. M. (1969). Consecutive intradimensional and extradimensional shifts in monkeys. *Journal of Comparative and Physiological Psychology*, *67*(2p1), 199.

Sutherland, N. S., & Mackintosh, N. J. (1971). Selective Attention and Learning.(Book Reviews: Mechanisms of Animal Discrimination Learning). *Science*, *173*, 715.

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.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & psychophysics*, *49*(1), 83-90.

Thomson, D. R., & Milliken, B. (2013). Revisiting the time course of inter-trial feature priming in singleton search. *Psychological research*, *77*(5), 637-650.

Treisman, A. M. (1964). Selective attention in man. British medical bulletin.

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

Turner, D. C., Aitken, M. R., Shanks, D. R., Sahakian, B. J., Robbins, T. W., Schwarzbauer, C., & Fletcher, P. C. (2004). The role of the lateral frontal cortex in causal associative learning: exploring preventative and super-learning. *Cerebral Cortex*, *14*(8), 872-880.

Vanni-Mercier, G., Mauguiere, F., Isnard, J., & Dreher, J. C. (2009). The hippocampus codes the uncertainty of cue–outcome associations: an intracranial electrophysiological study in humans. *Journal of Neuroscience*, *29*(16), 5287-5294.

Vogel, E. H., & Wagner, A. R. (2017). A theoretical note in interpretation of the "redundancy effect" in associative learning.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animals. In N. E. Miller & R. R. Spear (Eds.), Information processes in animals: Memory mechanisms (pp. 95–128). Hillsdale, NJ: Erlbaum.

Wagner, A. R., Brandon, S. E., Klein, S. B., & Mowrer, R. R. (1989). Evolution of a structured connectionist model of Pavlovian conditioning (AESOP). *Contemporary learning theories: Pavlovian conditioning and the status of traditional learning theory*, 149-189.

Watson, P., Pearson, D., Chow, M., Theeuwes, J., Wiers, R. W., Most, S. B., & Le Pelley, M. E. (2019). Capture and control: Working memory modulates attentional capture by reward-related stimuli. *Psychological science*, *30*(8), 1174-1185.

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.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, *1*(2), 202-238.