Protection from uncertainty in the exploration/exploitation trade-off

Adrian R. Walker[1], Danielle J. Navarro[1], Ben R. Newell[1], Tom Beesley[1, 2]

1. School of Psychology, UNSW Sydney, Australia
2. Department of Psychology, Lancaster University, UK

Address correspondence to:

Adrian R. Walker

School of Psychology, UNSW Sydney, NSW 2052, Australia

adrian.walker@unsw.edu.au

Raw data and analysis code can be found at:

https://osf.io/kra8p/?view_only=dde91e8ac50b4ed09c8c1fc333a735a0

Abstract

The exploration/exploitation trade-off (EE trade-off) describes how, when faced with several competing alternatives, decision-makers must often choose between a known good alternative (exploitation) and one or more unknown but potentially more rewarding alternatives (exploration). Prevailing theory on how humans perform the EE trade-off states that uncertainty is a major motivator for exploration: the more uncertain the environment, the more exploration that will occur. The current paper examines whether exploratory behaviour in both choice and attention may be impacted differently depending on whether uncertainty is onset suddenly (unexpected uncertainty), or more slowly (expected uncertainty). It is shown that when uncertainty was expected, participants tended to explore less with their choices, but not their attention, than when it was unexpected. Crucially, the impact of this "protection from uncertainty" on exploration only occurred when participants had an opportunity to learn the structure of the task prior to experiencing uncertainty. This suggests that the interaction between uncertainty and exploration is more nuanced than simply more uncertainty leading to more exploration, and that attention and choice behaviour may index separate aspects of the EE trade-off.

Keywords: Cognition, Decision Making, Reinforcement Learning, Attention, Exploration/Exploitation Trade-Off

Imagine that Mary lives in a large city and is driving to work. From previous experience she has learned that, on average, her trip to work takes approximately 30 minutes. However, the traffic is unpredictable, and in practice her commute can take anywhere between 15 to 45 minutes. For the last two days it has taken 45 minutes due to traffic — Mary attributes this to the normal variability and continues with her usual route. In contrast, Nick lives in a smaller town — his commute also takes 30 minutes on average but the traffic is highly predictable and the commute is always between 25 to 35 minutes. For the last two days, however, it has taken Nick 45 minutes to get to work due to traffic. He concludes that something has changed and decides to try a different route. The choices made by Nick and Mary both seem sensible due to the differences in the kind of uncertainty they face. In Mary's situation her uncertainty (about today's commute) is entirely due to the usual *expected* day to day variation in outcomes, and she continues to exploit her knowledge of the world (by following the route that has previously worked for her). In Nick's case, however, the uncertainty is *unexpected*: a previously predictable result has suddenly shown new and unexplained variability, prompting him to explore a different route.

The key point of this example is that Mary and Nick, though experiencing a similar delay to work, act differently depending on their expectations of the normal variability in the environment. When faced with uncertainty Mary (who expects a high level of uncertainty) continues to *exploit* her current known best route, while Nick (who does not expect a high level of uncertainty) *explores* for new routes. This idea was recently discussed by Cohen, McClure, and Yu (2007), who argued that the way that participants perform this *exploration/exploitation trade-off* (EE trade-off — the trade-off between choosing known good alternatives and unknown but potentially better alternatives) may change depending on whether decision-makers experience *unexpected uncertainty*, or *expected uncertainty*. The logic of this is that, when the environment suddenly appears to change (i.e., the environment becomes uncertain unexpectedly), it may indicate that it is necessary for a decision-maker to change their behaviour to adapt to the new environment. However, if the environment is generally stable with

some predictable variability (i.e., the environment is expected to be uncertain), then there is less reason to believe that there is new information to be gained by exploring.

The idea of conceptualising uncertainty into unexpected uncertainty and expected uncertainty has received growing support. Neurological research has shown that when uncertainty occurs unexpectedly, there is an increase in the speed of learning accompanied by a change in cortical activation, possibly in an attempt to learn what has changed in the environment. For example, it has been shown that the neuromodulator acetylcholine appears to index expected uncertainty, while norepinephrine, noradrenaline, and dopamine appear to index unexpected uncertainty (Marshall et al., 2016; Yu & Dayan, 2005), and different cortical regions are responsible for processing expected and unexpected uncertainty (Payzan-LeNestour, Dunne, Bossaerts, & O'Doherty, 2013). In the field of associative learning, Courville, Daw, and Touretzky (2006) have proposed that when learners experience unexpected change, they interpret a fundamental switch in the *latent cause* (the underlying rules) of the task. If a participant detects that the latent cause has changed, they should prepare themselves to learn any new rules to quickly adapt to the new environment. Following Courville et al.'s (2006) paper, the *latent cause theory* has been successfully applied to both simulating existing learning phenomena (Courville et al., 2006; Gershman, Blei, & Niv, 2010; Gershman & Niv, 2012) and motivating novel experimental work (Easdale, Le Pelley, & Beesley, 2019; Gershman, Jones, Norman, Monfils, & Niv, 2013).

If it is the case that unexpected uncertainty facilitates learning more than expected uncertainty, it is logical to assume that exploration should increase more under unexpected uncertainty than expected uncertainty. That is, if an agent thinks that the environment has changed, it should be aiming to explore as many actions as possible and observe their effects (maximising information gain while readiness to learn is high). By contrast, if an organism thinks that the environment has remained the same, it should be trying to exploit a small number of actions rather than expending effort learning about the environment (assuming the environment is already fairly well known).

The current paper explores this interaction between expected and unexpected uncertainty and the EE trade-off. Given the vast majority of everyday decisions are made under some level of uncertainty (i.e. where the outcomes expected from a given action are not well known, Mehlhorn et al., 2015), understanding what aspects of uncertainty affect decision-making is critical to understanding human decision-making more broadly. Specifically, the current paper aims to begin the process of empirically untangling uncertainty's interaction with the EE trade-off by examining how behaviour differs following expected uncertainty and unexpected uncertainty.

To test how expected and unexpected uncertainty and the EE trade-off may interact, we used two behavioural metrics: choice behaviour and attention. Choice behaviour (how participants allocate their choices between a series of alternatives) is the traditional method of assessing the EE trade-off, with selections of the best known alternatives considered to be exploitative, and selection of any other alternative to be exploratory (Mehlhorn et al., 2015). These choices are generally examined in *multi-armed bandit tasks* (Bradt, Johnson, & Karlin, 1956; Gittins, 1979). In the multi-armed bandit task, participants are presented a series of choices between several alternatives or *arms*, and, on each trial, are required to choose between them. The participants are then rewarded some number of points based on the arm that they picked on that trial. This simple task has been important in assessing the EE trade-off in humans. For example, Daw, O'Doherty, Dayan, Seymour, and Dolan (2006) found participants would preferentially explore arms that had been associated with high-value rewards in the past, and Speekenbrink and Konstantinidis (2015) found that participants would make more exploratory choices when rewards were highly variable.

As well as assessing how participants' allocate their choices, assessing how participants use their attention to solve EE trade-off problems has recently gained traction in the broader learning literature (Beesley, Nguyen, Pearson, & Le Pelley, 2015; Easdale et al., 2019; Walker, Le Pelley, & Beesley, 2017). The *attentional EE trade-off*, similar to the EE trade-off in choice, describes how decision-makers must often choose between attending to known useful information (that will help them make a decision),

and unknown but potentially more useful information (Beesley et al., 2015). To assess the EE trade-off in attention, the current paper employs a version of the *learned predictiveness design* (Le Pelley & McLaren, 2003). In the learned predictiveness design, participants are shown two cues, and asked to make one of two responses. In this task, the correct response differs from trial to trial, and depends on which cues are present. On every trial there are two cues, a *predictive cue* (cue `A` or `B`), and one *non-predictive cue* (cue `X` or `Y`). The predictive cues inform which response will be correct on that trial, while the non-predictive cues provide no information about the correct response on that trial. Crucially, once the participant learns the relationship between predictive cues and responses, they should be able to choose the correct response on every trial.

Previous research using the learned predictiveness design has showed that over time, participants begin to preferentially attend to predictive cues over non-predictive cues (Le Pelley, Beesley, & Griffiths, 2011; Le Pelley & McLaren, 2003). Beesley et al. (2015) have argued that this preference for predictive over non-predictive cues represents *attentional exploitation*, with participants pruning out the irrelevant non-predictive cue from processing (Niv et al., 2015; Rehder & Hoffman, 2005). Furthermore, they showed that when the validity of the predictive cue is reduced, such that it only predicts the correct response on two thirds of all trials, participants show increased attention to both predictive and non-predictive cues. Beesley et al. (2015) suggested that this may reflect *attentional exploration*, as participants look for information to help them reduce the uncertainty in the task.

The current paper combines the learned predictiveness design with the multi-armed bandit task to produce a *two-armed contextual bandit task* (Schulz, Konstantinidis, & Speekenbrink, 2018; Walker, Luque, Le Pelley, & Beesley, 2019). The contextual bandit task is similar to the traditional bandit task: on each trial there are several arms that the participant can pick, and when an arm is picked it pays out some reward value in points. However, the difference is that in a contextual bandit task the value of arms changes depending on the *context* for the decision, usually indicated by some visual cue.

To give a concrete example, in the current set of experiments the participants are tasked with selling a combination of two chemicals (represented as pictures of molecules) to one of two aliens that will pay them in alien currency for the chemicals. Importantly, only one of the chemicals on offer for sale determines the amount of currency each alien will pay. That is, one chemical is the *predictive cue* that determines the context, while the other chemical is the *non-predictive cue*. The two aliens are the *arms* that change in value depending on the context set-up by the predictive cue, and the participant's job is to learn which chemical cue predicts which alien they should sell to in order to earn the greatest rewards.

This two-armed contextual bandit task provides the ideal platform for examining the attentional EE trade-off. As only *one* of the cues is relevant for predicting the value of responses (the predictive cue), attention can be compared between the predictive and non-predictive cue to index an attentional EE trade-off. Furthermore, as the cues themselves do not have any value ascribed to them, and it is unnecessary to look at them in order to select alternatives in the task, attention to the cues should only index participants' attempts to explore them for information (and not index any attention required to physically click on an arm, Manohar & Husain, 2013).

This type of contextual bandit task has been used previously to examine the impact of uncertainty on choice behaviour and attention. Walker et al. (2019) showed that attentional exploration and exploitation appeared to co-occur with exploration and exploitation in choice behaviour. They postulated that, though exploratory behaviour co-occurred in both attention and choice behaviour, this represented two distinct processes. In choice, participants needed to explore to learn the *value* of rewards, while in attention they needed to explore to learn how well the cues *predicted* those rewards. Walker et al. (2019) found that when participants engaged in exploration under uncertainty with their choice behaviour, they also appeared to engage in exploration with their attention. Crucially, Walker et al. (2019) only assessed the impact of expected uncertainty on exploratory attention. Given that Cohen et al. (2007) postulated that unexpected uncertainty and expected uncertainty may have differing

impacts on exploratory choice, it is not unreasonable to think that this may also be the case for attention.

To summarise, the overall aim of the current study was to directly compare how participants perform the EE trade-off in both choice and attention when uncertainty was expected versus when it was unexpected. This was accomplished by employing a contextual two-armed bandit task, which made it possible to assess the EE trade-off in both choice and attention. Across four experiments, it is shown that there is a difference in exploration following unexpected and expected uncertainty, but only when participants have been given the opportunity to learn the best response strategy prior to experiencing expected uncertainty. However, this difference in exploratory behaviour following expected and unexpected uncertainty is not replicated in attention, with participants' exploratory attention appearing to be driven by the absolute level of uncertainty in the environment.

Experiment 1

Experiment 1 aimed to provide a comparison of the impact of unexpected uncertainty and expected uncertainty on the EE trade-off in choice and attention. By using a two-armed contextual bandit task in the style of the learned predictiveness design (Le Pelley & McLaren, 2003), it was possible to separate choice behaviour (measured by selection of responses) and attention (measured by attention to cues).

On each trial, participants were shown two chemical cues which determined the context of that trial. One of the chemical cues (the predictive cue) determined which one of two alien button responses would on average pay out the most points on that trial (the *high-value response*). The other chemical cue (the non-predictive cue) was task-irrelevant. There were two conditions in the experiment, the *sudden change* condition and the *always uncertain* condition. The condition determined how participants would experience uncertainty (operationalised as the level of variability in rewards). This was done by splitting the experiment into two stages. In the *sudden change* condition, the rewards associated with participants' choices were completely

deterministic during stage one, and as a consequence the uncertainty arose suddenly at the beginning of stage two. By contrast, in the *always uncertain* condition, the task began with a high level of uncertainty and stayed at that level for the entire task. This meant that participants in the *sudden change* condition would experience unexpected uncertainty, but participants in the *always uncertain* condition would experience expected uncertainty.

We indexed exploratory behaviour in choice as the amount of *low-value responses* (the response that on average yielded less reward) made during the task. In attention, an increased proportion of trial time attending to the predictive cue over the non-predictive cue was interpreted to indicate attentional exploitation, while an increased proportion of trial time attending to cues overall was interpreted to indicate attentional exploration. Because unexpected uncertainty has been theorised to motivate exploration more than expected uncertainty (Cohen et al., 2007), it was predicted that in stage two, participants in the *sudden change* condition would show more exploratory behaviour in both choice and attention than participants in the *always uncertain* condition.

Method

Participants. Sixty-one participants were recruited from UNSW Sydney in exchange for course credit. The mean age was 20.5 years (SD = 4.22); 43 participants identified as female and 18 as male. Testing continued until there were 24 participants in each condition that did not have to be excluded, leaving 48 participants for the final analysis. During testing, eight participants were excluded due to having fewer than 50% of trials with at least one fixation on a cue. Five participants who did not complete all the trials in the allocated time were also excluded. The two highest performing participants in each condition were paid $20 after data collection had finished.

Materials

The experiment was implemented in MATLAB using the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Participants' gaze was

tracked using a Tobii TX-300 eye-tracker (Tobii Technology, Danderyd, Sweden) connected to a 23-inch monitor (1920 by 1080 pixels) that sampled gaze location at 300 Hz. Participants were positioned in a chin-rest approximately 55 centimetres from the screen. As shown in Figure 1, cues were presented as cartoon depictions of four molecules, displayed on screen as 500 by 375 pixel images (visual angle 13.8° by 10.5°). The left cue was centred 384 pixels (10.6°) from left edge of the screen, while the right cue was centred 1536 pixels (40.7°) from the left edge of the screen. Both cues were shown 270 pixels (7.5°) from the top edge of the screen. The two response options were cartoon depictions of aliens, each of which was 200 pixels wide by 200 pixels tall (visual angle 5.6° by 5.6°). The upper response option was centred at 648 pixels (18.7°) from the top edge of the screen, and the bottom response option was centred 864 pixels (23.8°) from the top edge of the screen, with both response options centred horizontally. The four images used for cue stimuli were randomly allocated for each participant, and these allocations were yoked across conditions.

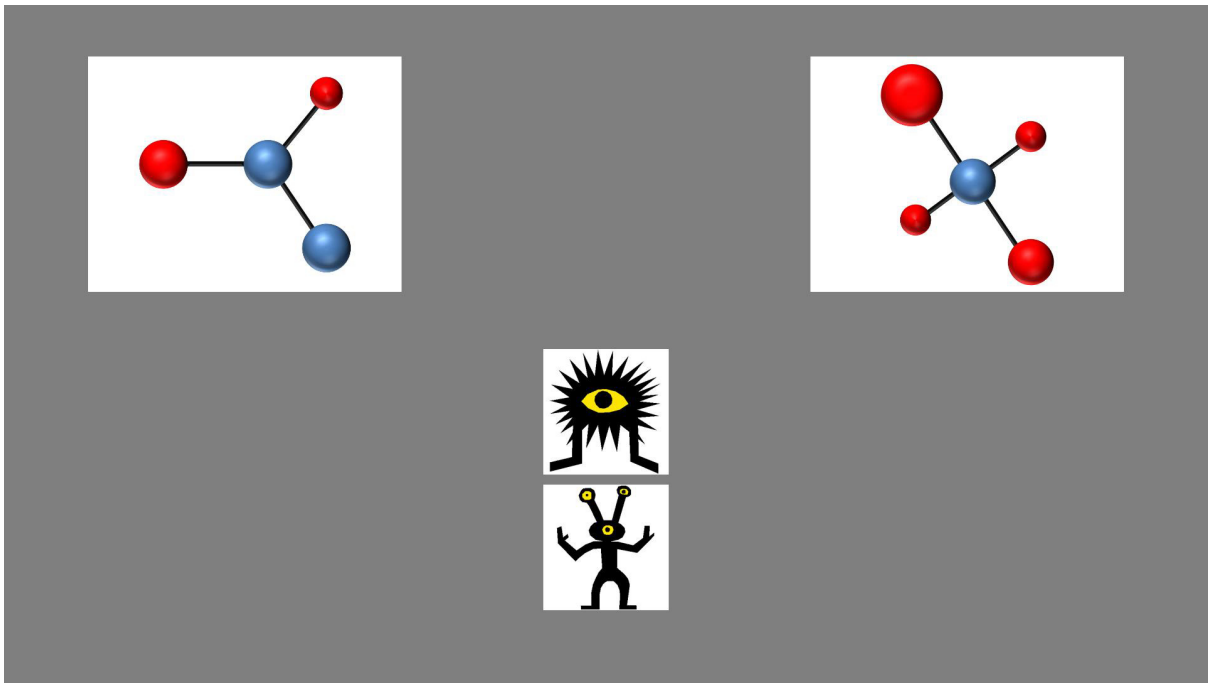

Figure 1. An example trial from Experiment 1. Participants were shown two molecules (cues), and told that they were selling a chemical mixture of these molecules to one of two aliens (responses). They could sell to an alien by clicking on it, and each alien would pay a certain amount of "sparflex" (an imaginary currency) for the mixture. Cue and response stimuli were taken from Beesley et al. (2015).

Design. The design of all conditions used in the current paper is shown in Table 1. Of the four cues in the design, two (cues `A` and `B`) were predictive cues, while the other two (cues `X` and `Y`) were non-predictive cues. The predictive cues were relevant to completing the task, while the non-predictive cues were task irrelevant. In all conditions, when cue `A` was present, response `1` (`R1`) would confer more points on average, while when cue `B` was present, response `2` (`R2`) would confer more points on average. Experiment 1 includes the first two conditions of Table 1, the *sudden change* condition, and the *always uncertain* condition. In the *sudden change* condition, the rewards were certain (i.e., did not vary) in stage one: the high-value response always resulted in a reward of 15 points and the low-value response always produced a reward of 10 points. In stage two, however, the reward for choosing the high-value response varied between 8 and 22 points (chosen randomly from a uniform distribution) and the reward for the low-value response varied from 3 to 17 points. This was represented in the form $U(9, 21)$ and $U(4, 16)$ for high-value and low-value response respectively. In the *always uncertain* condition, the rewards varied uniformly from 8-22 or 3-17 in both stages of the task. Crucially, the relationship between cues and responses never changed during the task, such that the high-value response always paid out the highest average reward in the two conditions.

Table 1
The cues and the associated reward outcomes for each response (`R1` and `R2`) for all four conditions used in the four experiments. In the sudden change *condition (in Experiments 1-4) there was no variability in rewards during stage one, while rewards varied by seven points either side of the mean value in the* always uncertain *condition (in Experiments 1 and 2). In the* gradual change *condition (in Experiments 2 and 3), rewards gradually became more variable during stage one until they varied six points either side of the mean score value. Finally, in the* mixed change *condition (in Experiments 3 and 4), rewards started without variability, and then shifted abruptly to a "peaked" distribution early during stage one. In stage two of all experiments, all rewards varied by seven points either side of the mean in all conditions with a uniform distribution.*

		Stage one		Stage two	
Condition	Cue pair	R1	R2	R1	R2
Sudden change	AX	15	10	$U(8, 22)$	$U(3, 17)$
	AY	15	10	$U(8, 22)$	$U(3, 17)$
	BX	10	15	$U(3, 17)$	$U(8, 22)$
	BY	10	15	$U(3, 17)$	$U(8, 22)$
Always uncertain	AX	$U(8, 22)$	$U(3, 17)$	$U(8, 22)$	$U(3, 17)$
	AY	$U(8, 22)$	$U(3, 17)$	$U(8, 22)$	$U(3, 17)$
	BX	$U(3, 17)$	$U(8, 22)$	$U(3, 17)$	$U(8, 22)$
	BY	$U(3, 17)$	$U(8, 22)$	$U(3, 17)$	$U(8, 22)$
Gradual change	AX	$15 \rightarrow U(9, 21)$	$10 \rightarrow U(4, 16)$	$U(8, 22)$	$U(3, 17)$
	AY	$15 \rightarrow U(9, 21)$	$10 \rightarrow U(4, 16)$	$U(8, 22)$	$U(3, 17)$
	BX	$10 \rightarrow U(4, 16)$	$10 \rightarrow U(9, 21)$	$U(3, 17)$	$U(8, 22)$
	BY	$10 \rightarrow U(4, 16)$	$10 \rightarrow U(9, 21)$	$U(3, 17)$	$U(8, 22)$
Mixed change	AX	$15 \rightarrow P(9, 21)$	$10 \rightarrow P(4, 16)$	$U(8, 22)$	$U(3, 17)$
	AY	$15 \rightarrow P(9, 21)$	$10 \rightarrow P(4, 16)$	$U(8, 22)$	$U(3, 17)$
	BX	$10 \rightarrow P(4, 16)$	$10 \rightarrow P(9, 21)$	$U(3, 17)$	$U(8, 22)$
	BY	$10 \rightarrow P(4, 16)$	$10 \rightarrow P(9, 21)$	$U(3, 17)$	$U(8, 22)$

Procedure. At the commencement of the experiment, participants completed a 7-point calibration of the eye-tracker, and were told that they must use the chin-rest while completing the experiment. Participants were told they would play the role of a salesperson trying to sell Earthen chemicals to aliens in exchange for a reward of "sparflex", a fictitious alien currency, and were asked to choose the chemicals presented in a way that would maximise their rewards. Before the experiment commenced, they were also informed that the top two performers in each condition (those who earned the highest numbers of points overall) would receive a $20 prize.

Stage one consisted of 256 trials divided evenly into 8 blocks of 32 trials, and stage two consisted of 256 trials divided evenly into 8 blocks of 32 trials. Every 64 trials, participants were given an self-paced rest break to reduce fatigue. In both stages, the cue pairings (`AX`, `AY`, `BX`, `BY`), positions of each cue (left or right), and positions of each response option (top or bottom) were counterbalanced, such that every 16 trials each possible combination of these factors was presented exactly once. The shift between the two stages was not signalled or announced to participants in any explicit way.

Each trial began with a black fixation cross presented in the middle of the screen for 1 second, followed by the presentation of the cue stimuli and the response options (Figure 1). Participants had unlimited time to view the stimuli before responding. Following the response, a feedback message appeared between the two cues and remained there for 2 seconds. The feedback message consisted of of the points received on that trial as well as the total points accumulated so far throughout the task. During this time, both response options remained on screen, and the selected response was outlined with a thick black border. Following the presentation of feedback, the next trial began immediately.

Every 32 trials, participants were probed for their knowledge of the relationship between cues and responses. On each probe trial, participants were shown each chemical cue (cues `A`, `B`, `X`, and `Y`) one at a time, and asked to indicate which of the two alien response buttons would pay the most sparflex for that chemical cue. After they had answered, they were probed for a confidence rating of their response from 1 to

5, with 1 representing 'I am guessing' and 5 representing 'I am certain'.[1]

Results

Data were split into blocks of 32 trials for analysis[2]. Trials that were longer than two standard deviations above or below the mean trial time were excluded from analysis. When this criterion was applied, a median of 2 trials and mean of 1.7 trials per block were removed. The key behavioural results are summarised for brevity in this section, but the full results of all tests can be seen in Supplementary B.

Response behaviour. Throughout the paper, choice data were analysed using three omnibus logit regressions, one in stage one, one in the transition between stages one and two (in the case of the current experiment, blocks 8 and 9), and one in stage two. The regressions were done in R with the lme4 package (Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2015), with random intercepts for each participant. The regressions included fixed effects of condition (with sum to zero contrasts) and block (with polynomial contrasts), with a random effect of participant, and were followed up by a Type 3 sums of squares ANOVA using the car package (Fox & Weisberg, 2011). Follow-up comparisons were using the Benjamini-Hochberg procedure (as in Konstantinidis, Taylor, & Newell, 2018), to adjust for the different stages of choice and attention separately.

The choice data from Experiment 1 can be seen in Figure 2. In stage one, participants in both conditions showed learning of the relationship between the predictive cue and the high-value response, $\chi^2(7) = 662.85, p < .001$, making more exploitative high-value responses over the course of stage one. Sensibly, this increase in exploitative responding was greater for participants in the *sudden change* condition, $\chi^2(3) = 221.77, p < .001$, with participants showing overall more exploratory (low-value) responses in the *always uncertain* condition, $\chi^2(1) = 104.31, p < .001$.

---

[1] These ratings were exploratory, and did not provide much insight into the cognition of the participants. In the interest of brevity, we do not present an analysis of the confidence data in this paper. The data itself can be seen in Supplementary A.

[2] Raw data for all experiments can be found at https://osf.io/kra8p/?view_only=dde91e8ac50b4ed09c8c1fc333a735a0

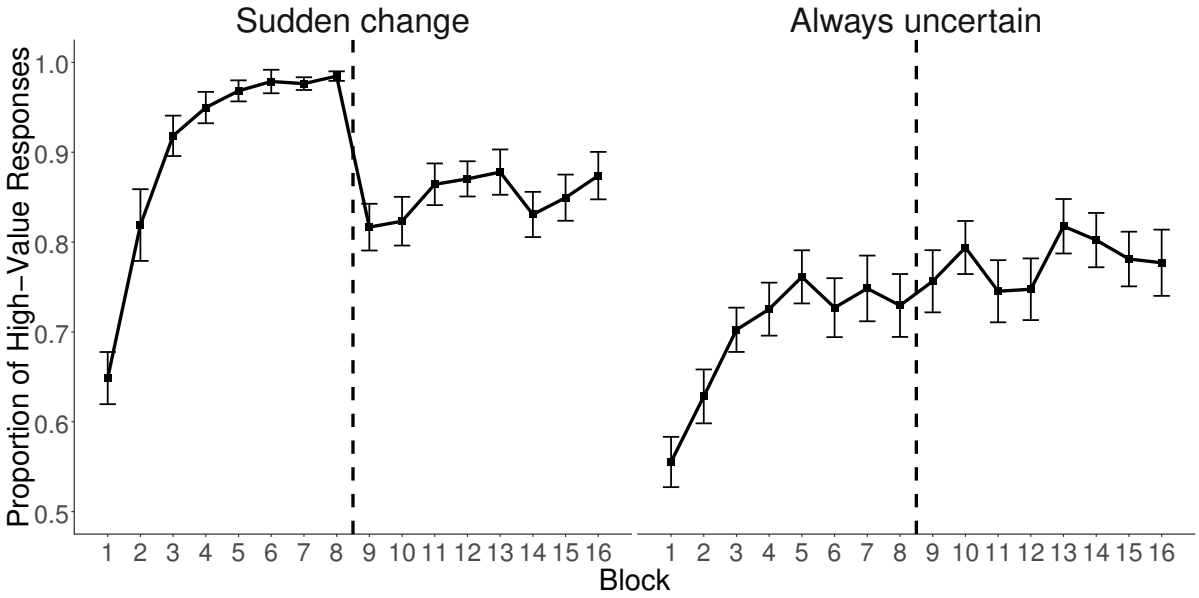

Figure 2. The choice data from both conditions of Experiment 1. Stage one occurred during blocks 1 through 8, and stage two occurred on blocks 9 through 16, with the dotted vertical line indicating where the stages changed. In the *sudden change* condition, rewards became more variable in stage two. In the *always uncertain* condition, stages one and two had a high level of variability. Error bars represent standard error of the mean.

In the transition from stage one to stage two, participants in the *sudden change* condition showed a decrease in selection of high-value responses that was not observed in the *always uncertain* condition, $\chi^2(1) = 72.73, p < .001$. Importantly, while participants in the *sudden change* condition made more high-value responses in block 8, $b = 3.35, z = 7.97, p < .001$, 95% PLCI $[2.57, 4.25]$, this difference between conditions disappeared in block 9, $b = 0.35, z = 1.25, p < .210$, 95% PLCI $[-0.22, 0.92]$. From this, it is clear that participants who experienced unexpected uncertainty in the *sudden change* condition showed a decrease in selecting high-value responses. However, there was no evidence that participants who experienced unexpected uncertainty were prompted to explore any more than participants who had experienced expected uncertainty. In stage two, participants in both conditions showed some improvement in selecting high-value responses, $\chi^2(7) = 22.89, p = .002$, though there was no significant difference between conditions, $\chi^2(1) = 3.62, p = .057$. Again, this suggested that unexpected uncertainty did not prompt more exploration compared to expected

uncertainty.

Attention. A dispersion-threshold identification algorithm (Salvucci & Goldberg, 2000) was used to process the eye-tracking data. A fixation was determined to have occurred when eye-gaze was contained within a maximum dispersion threshold of 75 pixels for at least 150 milliseconds. Fixation position was determined by the mean horizontal and vertical pixel values across the fixation sample. The eye that had the fewest missing samples on each trial was used for the analysis of that trial. Gaps in the data of no longer than 75 milliseconds were interpolated between the start of the data gap and the end of the data gap. Once these data were processed, the proportion of trial time each participant spent fixating on an area within each cue was calculated. Only attention to cues was relevant, with other time spent attending to response options or other space on the screen removed from analysis.

Throughout the paper, eye-gaze data were analysed using three omnibus linear regressions: one in stage one, one in the transition between stages one and two (in the case of the current experiment, blocks 8 and 9), and one in stage two. The regressions were done in R with the lme4 package (Bates et al., 2015; R Core Team, 2015), with random intercepts for each participant. The regressions included fixed effects of condition, cue predictiveness (both with sum to zero contrasts), and block (with polynomial contrasts), with a random effect of subject, and were followed up by a Type 3 sums of squares ANOVA using the car package (Fox & Weisberg, 2011). Follow-up comparisons were done adjusting for multiple comparisons using the Benjamini-Hochberg procedure (as in Konstantinidis et al., 2018), adjusting for the different stages of choice and attention separately.

The eye-gaze data from Experiment 1 can be seen in Figure 3. In stage one, participants showed a clear bias towards the predictive cue over the non predictive cue, $\chi^2(1) = 127.15, p < .001$, indicating attentional exploitation. This difference appeared to be greater in the *sudden change* condition compared to the *always uncertain* condition, $\chi^2(1) = 40.19, p < .001$, with participants who experienced no uncertainty better able to exploit their knowledge of the relationship between cues and responses.

This difference appeared to be driven entirely by a difference in attending to the non-predictive cue, with participants in the *sudden change* condition attending to the non-predictive cue significantly less than participants in the *always uncertain* condition, $b = -0.02, t(46) = -3.09, p = .007$, 95% PLCI $[-.02, -0.01]$.

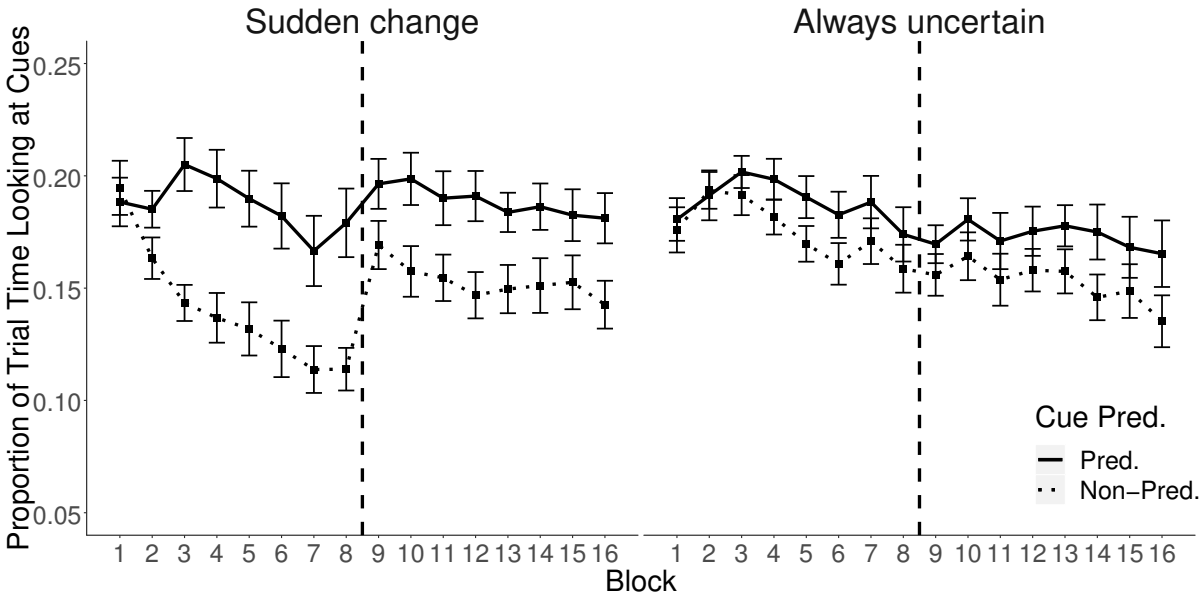

Figure 3. The eye-gaze data from both conditions of Experiment 1. Stage one occurred during blocks 1 through 8, and stage two occurred on blocks 9 through 16, with the dotted vertical line indicating where the stages changed. In the *sudden change* condition, rewards became more variable in stage two. In the *always uncertain* condition, stages one and two had a high level of variability. Error bars represent standard error of the mean.

In the transition from stage one to stage two, participants in the *sudden change* condition showed a significant increase in the proportion of trial time they spent attending to cues, $b = 0.04, t(71) = 3.69, p < .001$, 95% PLCI $[.02, 0.06]$, which was not observed in the *always uncertain* condition, $b = -3.56e-3, t(71) = -0.68, p = .501$, 95% PLCI $[-.01, 0.01]$. That is, it appeared that participants in the *sudden change* condition showed a re-engagement of attention to cues at the onset of stage two. However there was no difference in the proportion of trial time attending to cues between the two conditions in block 9, $b = 0.01, t(46) = 1.54, p = .131$, 95% PLCI $[-2.73e-3, 2.28e-2]$, with participants showing roughly equal levels of attentional exploration. It appeared that participants in the *sudden change* condition increased

their attention to the predictive cue more than the non-predictive cue compared to participants in the *always uncertain* condition. However, the three-way interaction was not significant, $\chi^2(1) = 3.69, p = .055$.

In stage two, there was no significant difference between conditions, $\chi^2(1) = 0.41, p = .524$. There was a significant interaction between condition and cue predictiveness, $\chi^2(1) = 11.25, p < .001$, with participants in the *sudden change* condition showing a greater difference in the proportion of trial time attending to the two cues compared to participants in the *always uncertain* condition. That is, surpisingly participants in the *sudden change* condition showed a greater level of attentional exploitation than participants in the *always uncertain* condition in stage two, despite receiving the same level of environmental uncertainty. Overall, it appears that participants who experienced unexpected uncertainty showed *less* exploratory behaviour than participants who had experienced expected uncertainty.

Discussion

Experiment 1 aimed to examine the difference between unexpected uncertainty and expected uncertainty on exploratory behaviour in choice and attention. From previous research, it was expected that when participants experienced unexpected uncertainty, they would show a greater level of exploratory behaviour than participants who had experienced expected uncertainty. This was because unexpected uncertainty would indicate a change in the structure of the environment that required new learning, driving exploration of the environment.

The results of Experiment 1 suggest that the opposite pattern may in fact be true. When participants experienced unexpected uncertainty, they did not appear to demonstrate any difference in exploration of their choice compared to participants who had previously experienced uncertainty. In fact, participants who experienced unexpected uncertainty showed a greater level of attentional exploitation following the onset of uncertainty, with a greater bias towards predictive information in the environment over non-predictive information in eye-gaze.

One reason this pattern of results may have emerged is the difference between learning of the relationship between cues and responses in each condition during stage one. In the *sudden change* condition, participants were given ample opportunity to learn the relationship between cues and responses in stage one. By contrast, this did not appear to occur for participants in the *always uncertain* condition, with participants struggling to exploit the high-value response in stage one, and only showing a small bias in attending to the predictive cue. As the relationship between cues and responses did not change between stages one and two, it is possible that participants in the *sudden change* condition were better able to transfer their knowledge from stage one to stage two. That is, they were able to maintain a clear bias in attending to the predictive cue even under uncertainty. It may be the case that to tease apart unexpected and expected uncertainty, it is necessary to match participants more closely on their knowledge of the task before experiencing the uncertainty in stage two. This idea is explored in Experiment 2.

Experiment 2

Experiment 2 aimed to examine the difference in unexpected uncertainty and expected uncertainty when participants' knowledge of the relationship between cues and responses was more closely matched between the two conditions than in Experiment 1. In Experiment 1, the key comparison was between the *sudden change* condition, where participants experienced no uncertainty (in the form of reward variability) followed by a high level of uncertainty, and the *always uncertain* condition, where participants always experienced a high level of uncertainty. In Experiment 2, a third condition was included: the *gradual change* condition. In the *gradual change* condition, participants began in stage one experiencing no uncertainty, but over the course of stage one uncertainty increased slowly leading into stage two. The purpose of this was to give participants a better chance to learn the relationship between cues and responses in stage one (as in the *sudden change* condition), but also not experience a sudden increase in uncertainty at the onset of stage two (as in the *always uncertain* condition).

It was predicted that, as participants in the *gradual change* condition would not experience unexpected uncertainty at the onset of stage two (but crucially should also know the relationship between cues and responses), they would show less exploratory behaviour in stage two than participants in either the *sudden change* or the *always uncertain* conditions. Unlike in Experiment 1, only choice data were collected in Experiment 2. This allowed a large number of participants to be gathered quickly, and establish whether it was the case that exploratory behaviour in choice differed in stage two between the *gradual change* condition and the other two conditions.

Method

Participants. One hundred and twenty six students from UNSW Sydney participated for course credit ($n = 93$), or for $16 cash payment ($n = 33$). The mean age was 20.4 years (SD = 3.61); 73 participants identified as female and 53 as male. To try and ensure that all participants learned something of the contingency between the cues and the rewards, a decision was made after the initial round of data collection (the first 105 participants) to exclude participants who performed below chance during the experiment (fewer than 50% of their choices high-value responses). In total, 19 participants were excluded due to failing to meet this criterion. Two more participants were excluded for failing to complete the task during the one hour time allocated for the task. Testing continued until there were 35 participants in each group that did not have to be excluded, leaving 105 participants for the final analysis. The two highest performing participants in each condition were paid $20 after data collection had finished.

Materials. All materials were the same as Experiment 1, with the exception that monitors were not connected to a Tobii eye-tracker, and participants were not required to put their chin in a chin-rest.

Design. The design of all three conditions is shown in Table 1. The only difference between Experiment 1 and 2 was that in Experiment 2 the *gradual change* condition was also included. In the *gradual change* condition, the level of uncertainty in

rewards was increased across the course of stage one. At the beginning of stage one, the rewards were deterministic and followed the same rule as for the *sudden change* condition: the high-value response elicited a reward of 15 and the low-value response elicited a reward of 10. After a brief period of certainty, the responses became variable. At first the rewards varied by one point on either side of the mean (e.g. a reward distribution of $U(14, 16)$ for a high-value response), with the range of possible rewards increasing linearly each block until the scores varied by 6 points either side of their mean (i.e., $U(9, 21)$ and $U(4, 16)$ for high-value and low-value response respectively). Thus, at the end of stage one (block 8) the rewards in the *gradual change* condition were almost as variable as in stage two. The key variable of interest was the amount of high-value responses participants made during the task, with more high-value responses indicating greater exploitation, and fewer high-value responses indicating greater exploration.

Procedure. The procedure was the same as that in Experiment 1, with the following changes: stage one consisted of 256 trials divided evenly into 8 blocks of 32 trials, and stage two consisted of 192 trials divided evenly into 6 blocks of 32 trials. Every 64 trials, participants were given an self-paced rest break to reduce fatigue. Unlike Experiment 1, participants were not probed for their knowledge of the relationships between cues and rewards.

Results

Data were split into blocks of 32 trials for analysis. Trials that were longer than two standard deviations above or below the mean trial time were excluded from analysis. When this criterion was applied, a median of 2 trials and mean of 1.7 trials per block were removed. The key behavioural results are summarised for brevity in this section, but the full results of all tests can be seen in Supplementary C.

The learning curves for all three conditions are plotted in Figure 4. Visual inspection of Figure 4 suggests participants learned the contingency in all three conditions, and this was corroborated in the regression analyses, $\chi^2(7) = 1191.47, p < .001$. Reflecting the fact that some conditions had less variability

than others during stage one, participants in the *sudden change* condition adopted the high-value response fastest, followed by participants in the *gradual change* condition and *always uncertain* conditions respectively, $\chi^2(14) = 429.35, p < .001$.

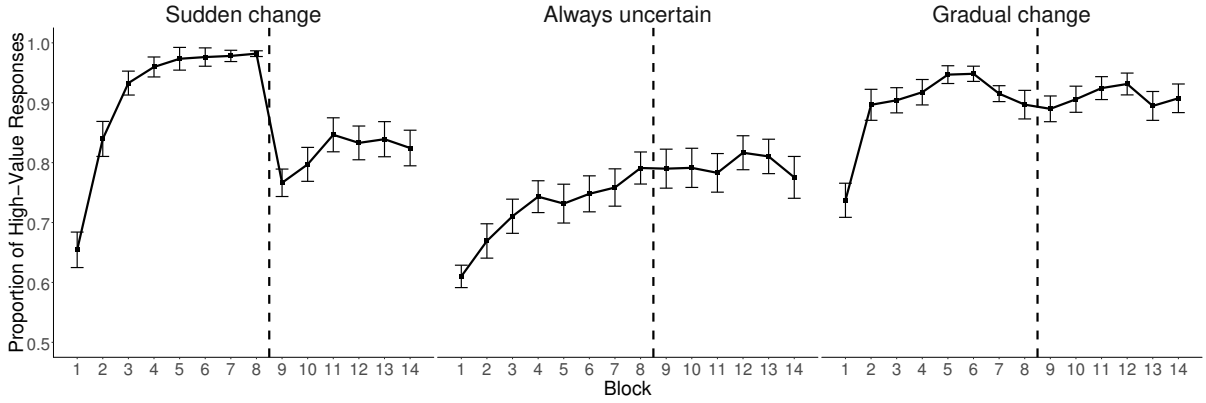

Figure 4. The choice data from all three conditions of Experiment 2. Stage one occurred during blocks 1 through 8, and stage two occurred on blocks 9 through 14, with the dotted vertical line indicating where the stages changed. In the *sudden change* condition, rewards became more variable in stage two. In the *always uncertain* condition, stages one and two had a high level of variability. In the *gradual change* condition, rewards became more variable slowly over the course of stage one leading into stage two. Error bars represent standard error of the mean.

During stage one, participants in the *always uncertain* condition made fewer high-value responses than participants in either the *sudden change*, $b = 1.55, z = 8.07, p < .001$, 95% PLCI $[1.17, 1.94]$, or the *gradual change* conditions, $b = 1.34, z = 7.10, p < .001$, 95% PLCI $[0.96, 1.72]$, suggesting that participants engaged in more exploratory behaviour when uncertainty was higher. An overall difference between the *sudden change* condition and *gradual change* condition was not detected $b = 0.22, z = 1.09, p = .370$, 95% PLCI $[-0.20, 0.64]$.

By the end of Stage 1, the three conditions were showing above chance selection of the high-value response, but as illustrated in Figure 4 they showed different levels of high-value responding at the start of stage two. At block 8, participants in the *sudden change* condition made more high-value responses than participants in the *gradual change* condition, $b = 1.82, z = 3.62, p = .001$, 95% PLCI $[0.85, 2.88]$, and the *always uncertain* condition, $b = 1.26, z = 3.45, p < .001$, 95% PLCI $[0.56, 2.01]$.

However, at block 9 participants in the *sudden change* condition made fewer high-value responses than participants in the *gradual change* condition, $b = 1.12, z = 4.18, p < .001$, 95% PLCI $[0.60, 1.67]$, suggesting that the onset of uncertainty produced an increase in exploratory behaviour. Moreover, this difference appears to persist beyond the first block of trials in stage two. Participants in the *gradual change* condition continued to make more high-value responses than participants in the *sudden change* condition across stage two as a whole, $b = 1.04, z = 3.73, p < .001$, 95% PLCI $[0.39, 1.60]$. Participants in the *gradual change* condition also maintained a higher level of performance than participants in the *always uncertain* condition, $b = 1.04, z = 3.18, p = .002$, 95% PLCI $[0.49, 1.69]$.[3]

Discussion

Experiment 2 aimed to explore the effect of unexpected and expected uncertainty on the EE trade-off when participants had ample opportunity to learn the relationship between cues and responses before transitioning to an uncertain environment. In Experiment 1, it was clear that when participants were not given a chance to learn the relationship between cues and responses prior to experiencing uncertainty (as in the *always uncertain* condition), they struggled to consistently exploit the high-value response. In order to elucidate the relationship between expected uncertainty and exploration, in Experiment 2 we included the *gradual change* condition, where initially participants experienced no uncertainty, before slowly transitioning into a high level of uncertainty. Crucially, while a sudden shift to uncertainty should have motivated exploratory behaviour (as a sudden switch to uncertainty may indicate a broader change in the environment), a slow transition into uncertainty should not.

Indeed, this intuition proved to be the case in the results. While participants in the *sudden change* and *gradual change* both performed close to optimally during stage

[3] For completeness: there were no significant differences in responding between the *sudden change* and *always uncertain* conditions, $b = 0.02, z = 0.06, p = .95$, 95% PLCI $[-0.57, 0.61]$. There was some improvement in stage two for the *sudden change* condition, $\chi^2(10) = 30.74$, $p < .001$, but nevertheless by block 14 there was still a significant difference between the *sudden change* and *gradual change* conditions , $b = 1.00, z = 2.57, p = .019$, 95% PLCI $[0.22, 1.82]$.

one, only those participants in the *gradual change* condition maintained that strategy of responding in stage two. This difference in performance was surprisingly persistent: over the course of stage two, participants in the *gradual change* condition continued to outperform participants in the *sudden change* condition long after one might have expected any difference to have disappeared. That is, participants who were gradually exposed to uncertainty appeared to have been *protected* from the effects of uncertainty on exploration, and as such we coin this effect the *protection from uncertainty* effect. In contrast, participants in the *always uncertain* condition maintained an exploratory pattern of low-value responding throughout stages one and two.

The preliminary conclusions from this are twofold. The first is that unexpected uncertainty seems to induce exploratory behaviour to a greater extent than expected uncertainty. When participants experience a gradual transition into uncertainty, they continue to exploit their knowledge of the learnt cue-response associations, rather than make exploratory response choices.

The second preliminary conclusion is that when participants are given an opportunity to learn the high-value response in a stable (low uncertainty) environment, they tend to continue to perform the high-value response in a high uncertainty environment if there is a gradual transition between the high and low uncertainty environments. Participants in the *always uncertain* condition in Experiments 1 and 2 struggled to learn to exploit the relationship between cues and responses in stage one, and consequently continued to show a high level of exploratory choice in stage two (performing at around the same level as participants in the *sudden change* condition). Crucially, this implies that simply experiencing uncertainty is not enough to protect participants from exploring, and that the gradual transition into uncertainty is necessary to reduce exploratory behaviour.

One other facet of the data was that participants in the *gradual change* condition showed an increase in selection of high-value responses up until block 6, after which there was an inflection point where selection of high-value responses started to decrease. This is likely due to the fact that at block 7, participants could receive a

reward for selecting the high-value response that was equal to the initial mean reward value of the low-value response. That is, participants could select a high-value response (with a mean reward of 15) and receive a score of 10 (the mean reward of the low-value response). If participants encode scores of 10 or less as low-value rewards, this may have influenced them to occasionally gamble on low-value responses in an attempt to earn high rewards (somewhat like probability matching, Shanks, Tunney, & McCarthy, 2002).

Experiment 3

Experiment 3 aimed to extend the findings of Experiment 2 in three ways. First, as Experiment 2 was an exploratory study, it was necessary to run a confirmatory study to demonstrate the robustness of this protection from uncertainty effect.[4]

Second, though it was demonstrated that the gradual onset of uncertainty could protect participants from switching away from the high-value response, it was unclear whether the gradual onset of uncertainty, or the mere presence of a moderate level of uncertainty in stage one caused this protection (Gureckis & Love, 2009). That is, though it was clear that participants in the *always uncertain* condition in Experiment 2 made more low-value responses in stage two than participants in the *gradual change* condition, they experienced a much higher level of uncertainty overall in stage one. Therefore, it is unclear whether it is the gradual move from certainty to uncertainty in stage one that was responsible for reducing exploratory responding in stage two, or the presence of a moderate level of uncertainty in stage two (compared to the high level in the *always uncertain* condition).

To test this, a new condition, the *mixed change* condition, was included. This condition had the same level of score variability as the *gradual change* condition in Experiment 2, however this variability was intermixed throughout stage one (rather than gradually introduced over the course of stage one). The consequence of this was

[4] The preregistration for this experiment can be found at https://osf.io/5b9yd/. The analysis was changed from ANOVA to regression, though this did not affect the conclusions drawn from the data. The results from the pre-registered ANOVA can be seen in the OSF repository, and any inconsistencies in the analyses noted.

that the gradual transition from stage one to stage two did not occur, but participants in this condition experienced the same level of uncertainty in stage one as participants in the *gradual change* condition.

Finally, Experiment 3 examined the role of attention in the protection from uncertainty effect. Specifically, Experiment 3 examined whether the gradual onset of uncertainty would increase any bias in attentoin towards the predictive cue in stage two, indicating an increased level of attentional exploitation. As Experiment 3 was primarily concerned with examining behaviour related to the protection from uncertainty effect, the *always uncertain* condition was omitted in Experiment 3.

For the *sudden change* and *gradual change* conditions, it was predicted that the pattern of responding would replicate that seen in Experiment 3. Mainly, that participants would make more exploitative high-values responses in stage two in the *gradual change* condition compared to the other two conditions. Furthermore, it was predicted that because the gradual onset of uncertainty was not present in the *mixed change* condition, participants in this condition would make fewer high-value responses than participants in the *sudden change* and *gradual change* conditions in stage one, and subsequently fewer high-value responses than the *gradual change* condition in stage two (as these participants would be protected from the uncertainty in stage two). That is, we expected the mixed condition would not offer the same "protection from uncertainty" as the gradual condition. Finally, it was predicted that overt attention would align with response behaviour. As in Experiment 1, exploitative attention was indexed by a bias of attending towards the predictive cue over the non-predictive cue, while exploratory attention was indexed by the total level of attention to cues. Thus, it was predicted that when participants showed more exploratory responding, they would have a more exploratory pattern of attention, and the same should occur for exploitative responding.

Method

Participants. One hundred and nineteen students from UNSW Sydney participated for course credit ($n = 78$), or for \$16 cash payment ($n = 41$). The mean

age was 22.1 years (SD = 6.04); 70 participants identified as female and 49 as male. Exclusion criteria were preregistered to exclude participants with an average of under 55% high-value responding during stage one. This cut-off criterion was increased from Experiment 2, where it was set at 50% (the strictest cut-off to determine below-chance performance). The increase from a 50% cut-off to a 55% cut-off was done to ensure that participants who performed numerically slightly above chance, but were still in reality selecting arms randomly, were excluded from analysis. The decision to restrict exclusion to stage one was taken as to ensure that participants who increased their exploratory behaviour dramatically in stage two following uncertainty were not unreasonably excluded. 10 participants were excluded due to this criterion.

A further five participants were excluded due to having fewer than 50% of trials with at least one fixation on a cue, and six more were excluded for failing to complete the task during the one hour time allocated for the task. Testing continued until there were 32 participants in each condition who did not have to be excluded, leaving 96 participants for the final analysis. The two highest performing participants in each condition were paid $20 after data collection had finished.

Materials. In Experiment 3, participants' gaze was tracked using a Tobii TX-300 eye-tracker (Tobii Technology, Danderyd, Sweden) in the same fashion as in Experiment 1.

Design. Experiment 3 included three conditions: the *sudden change* condition and the *gradual change* condition (operationalised as in Experiment 2), and the *mixed change* condition. The *mixed change* condition experienced the same amount of uncertainty as the *gradual change* condition across stage one, but crucially did not experience the graduated increase in uncertainty. To do this, the rewards for stage one of the *mixed change* condition were first generated in the same fashion as those in the *gradual change* condition. After generation, the rewards for all the trials in blocks 3 to 8 were shuffled. Crucially, this meant that in every block in stage one after the introduction of uncertainty, participants could experience the full variability in the scores (from 6 points above the mean to 6 points below). It is important to note

however that the resulting distributions of scores in each block were not uniform. Instead, they emulated the shape of truncated normal distributions, with centres at 15 and 10 points (the mean value of high-value and low-value responses), and tails truncated at 6 points above and 6 points below the means. Therefore, participants in this condition experienced many scores close to the mean, and few scores far from the mean in stage one. This distribution of scores is referred to as $P(9, 21)$ for the high-value response and $P(4, 16)$ for the low-value response (P standing for "Peaked"). The design of the new *mixed change* condition can be seen in Table 1.

As in Experiments 1 and 2, the proportion of high-value responses participants made in the task was measured, with fewer high-value responses indicating greater exploration. Participants' gaze was measured in the same fashion as Experiment 1. Exploitative attention was operationalised by the proportion of trial time the participant spent looking at the predictive cue over the non-predictive cue, while exploratory attention was operationalised by the total proportion of trial time looking at both cues.

Procedure. The procedure was the same as in Experiment 2, with a few key exceptions. At the commencement of the experiment, participants completed a 7-point calibration of the eye-tracker, and were told that they must use the chin-rest while completing the experiment. In both the *gradual change* and *mixed change* conditions, uncertainty was introduced in block 3, with no variability in the rewards received in blocks 1 and 2 for any condition. To prevent participants from associating the onset of uncertainty with the presence of a rest break, in Experiments 3 and 4 rest breaks occurred every 100 trials, rather than every 64.

Results

Data were split into blocks of 32 trials for analysis. Trials that were longer than two standard deviations above or below the mean trial time were excluded from analysis. When this criterion was applied, a median of 2 trials and mean of 1.7 trials per block were removed. The key behavioural results are summarised for brevity in this

section, but the full results of all tests can be seen in Supplementary D.

Response behaviour. Response data from Experiment 3 can be seen in Figure 5. Given that the first two blocks of the experiment were identical in the three conditions, only blocks 3 to 8 were analysed in stage one. This was to ensure that the focus of the analysis would be on the effect of uncertainty on performance in stage one, without being influenced by the period of certainty in blocks 1 and 2[5].

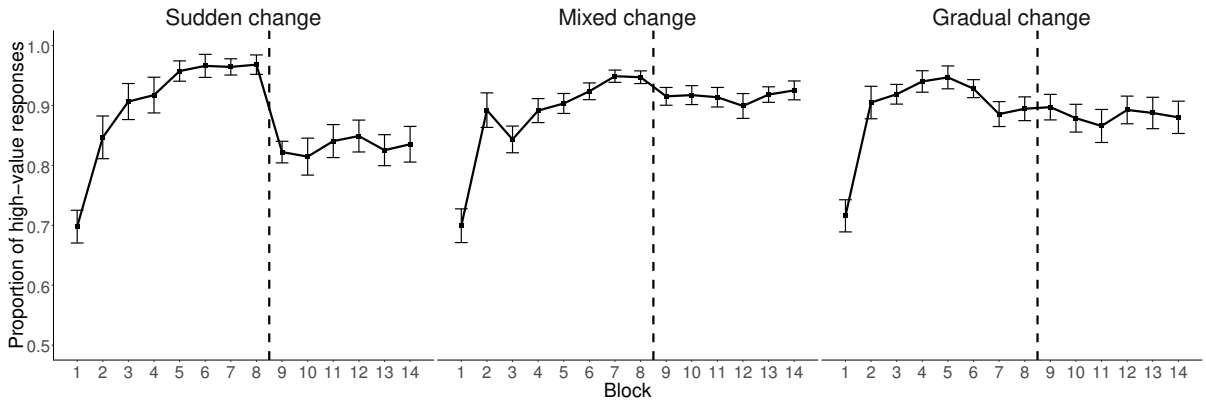

Figure 5. The choice data from all three conditions of Experiment 3. Stage one occurred during blocks 1 through 8, and stage two occurred on blocks 9 through 14, with the dotted vertical line indicating where the stages changed. In the *sudden change* condition, rewards became more variable in stage two. In the *mixed change* condition, rewards became moderately variable in block three of stage one, then more variable in stage two. In the *gradual change* condition, rewards became more variable slowly over the course of stage one leading into stage two. Error bars represent standard error of the mean.

The results of Experiment 3 show that the base protection from uncertainty effect for responding in the *gradual change* condition was partially replicated. In stage one, participants in the *sudden change* condition made more high-value responses than participants in the *gradual change* condition, $b = 0.81, z = 3.14, p = .003$, 95% PLCI $[0.32, 1.34]$. When transitioning from stage one to stage two, participants in the *sudden change* condition went from making significantly more high-value responses than participants in the *gradual change* condition in block 8, $b = 1.78, z = 4.00, p < .001$, 95% PLCI $[0.95, 2.73]$, to making significantly fewer high-value responses in block 9, $b = 0.90, z = 3.11, p = .003$, 95% PLCI $[0.35, 1.50]$.

[5] Cursory analysis of blocks 1 and 2 revealed no significant effect of condition, $\chi^2(2) = 1.92, p = .379$, suggesting that the three conditions were fairly well matched on performance during blocks 1 and 2.

However, contrary to the findings of Experiment 2, there was no significant difference between these two conditions overall during stage two, once corrected for multiple comparisons, $b = 0.64, z = 2.03, p = .064$, 95% PLCI $[0.01, 1.27]$. This is particularly surprising given that there was no suggestion that participants in the *sudden change* condition improved after the onset of uncertainty in stage two, nor that participants in the *gradual change* conditioned decreased in their level of high-value responding over stage two, $\chi^2(10) = 16.10, p = .097$.

Surprisingly, participants in the *mixed change* condition were able to learn to perform the task well during stage one, contrary to our predictions. Despite not receiving the gradual move from a certain environment to an uncertain environment, participants were able to learn to make the high-value response on nearly 100% of trials by the end of stage one, with no significant difference in performance compared to the participants in the *gradual change* condition in stage one, $b = 0.26, z = 1.19, p = .279$, 95% PLCI $[-0.17, 0.69]$. By block 8, participants in the *mixed change* condition were actually outperforming participants in the *gradual change* condition, $b = 0.76, z = 2.37, p = .027$, 95% PLCI $[0.13, 1.42]$, presumably due to the increased level of uncertainty in the *gradual change* condition, though participants in the *mixed change* condition still performed worse than participants in the *sudden change* condition, $b = 0.97, z = 2.15, p = .040$, 95% PLCI $[0.11, 1.92]$.

Transitioning into block 9, participants in the *mixed change* condition made significantly more high-value responses than participants in the *sudden change* condition, $b = 0.94, z = 4.05, p < .001$, 95% PLCI $[0.49, 1.43]$. Crucially, this performance persisted into stage two, with participants in the *mixed change* condition performing significantly better than participants in the *sudden change* condition in stage two, $b = 0.79, z = 2.94, p = .006$, 95% PLCI $[0.25, 1.33]$, demonstrating the protection from uncertainty effect. There was no evidence for a difference in the rate of high-value responding between the *mixed change* condition and the *gradual change* condition in stage two, $b = 0.16, z = 0.53, p = .597$, 95% PLCI $[-0.44, 0.75]$.

Attention. Eye-gaze data for Experiment 3 can be seen in Figure 6. As in the analysis of responses, only blocks 3 to 8 were analysed in stage one.[6]

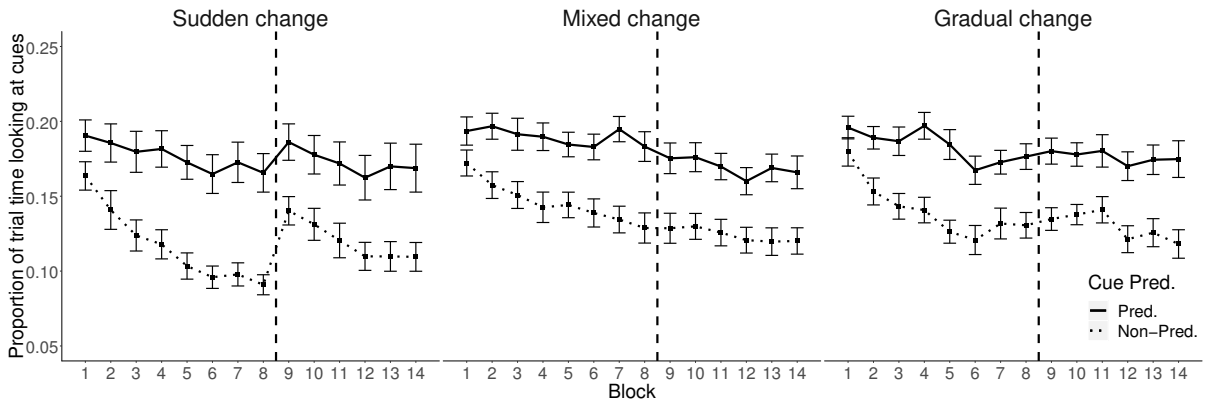

Figure 6. The eye-gaze data from all three conditions of Experiment 3. Stage one occurred during blocks 1 through 8, and stage two occurred on blocks 9 through 14, with the dotted vertical line indicating where the stages changed. In the *sudden change* condition, rewards became more variable in stage two. In the *mixed change* condition, rewards became moderately variable in block three of stage one, then more variable in stage two. In the *gradual change* condition, rewards became more variable slowly over the course of stage one leading into stage two. Error bars represent standard error of the mean.

Stage one showed some evidence for an attentional EE trade-off. Participants in the *sudden change* condition had a lower proportion of trial time looking at the non-predictive cue in stage one compared to participants in the *gradual change* condition, $b = -0.01, t(62) = -2.60, p = .035$, 95% PLCI $[-2.39e-2, -3.35e-3]$, and *mixed change* condition, $-b = 0.02, t(62) = -3.32, p = .009$, 95% PLCI $[-0.03, -0.01]$, though there was no difference observed between the conditions in the proportion of trial time looking at the predictive cue, ts $\leq 1.04, p$s $> .571$. This suggests participants in the *sudden change* condition were better able to exploit their knowledge of the contingencies to prioritise their attention to the predictive cue than the other two conditions.

It was also expected that participants in the uncertain *mixed change* and *gradual change* conditions would spend a greater proportion of trial time looking at cues in stage one in an attempt to explore them for information. However, though there was

[6] Cursory analysis of blocks 1 and 2 revealed no significant differences between any of the conditions in the proportion of trial time looking at cues, $\chi^2(2) = 0.94, p = .624$.

an overall effect of condition on the total proportion of trial time looking at cues in stage one, $\chi^2(2) = 6.23, p = .044$, follow-up tests did not suggest any differences between the conditions when corrected for multiple comparisons, ts $\leq 2.31, p$s $> .072$. As such, it is difficult to argue that participants spent longer exploring the cues in the *mixed change* and *gradual change* conditions, rather than simply distributing their attention to the predictive and non-predictive cues more equally than participants in the *sudden change* condition. This may be due to the fact that, unlike in previous research on the attentional EE trade-off (e.g. Beesley et al., 2015; Easdale et al., 2019; Walker et al., 2017), participants were generally able to perform the task well in all conditions (above 85% throughout blocks 3 to 8). As such, participants may have been less motivated to explore the cues for information to help them further solve the task.

Most surprisingly, there were no significant differences between the conditions in the transition between stage one and stage two once corrected for multiple comparisons, nor during stage two. Though there was a significant interaction between condition and block during the transition into uncertainty $\chi^2(2) = 22.04, p < .001$, follow-up contrasts between all conditions at blocks 8 and 9 did not show any significant differences, ts $\leq 2.25, p$s $> .099$. Furthermore, despite a difference between conditions in high-value responding during stage two, a difference between conditions in attention was not observed during stage two, $\chi^2(2) = 0.37, p = .831$.

The lack of a difference between conditions was particularly surprising, given the effect of condition observed in stage two of the response data. That is, while participants seem to be exploring more with their responses in the *sudden change* condition than the *mixed change* condition, this did not appear to be the case in looking time. To assess the evidence against a difference between the conditions in stage two, a Bayesian ANOVA using the BayesFactor package in R (Morey & Rouder, 2015) was run, with fixed factors of condition, block, and cue predictiveness, and a random factor of participant. Two million iterations of the generating Monte-Carlo algorithm were run, assuming a flat prior which weighted all possible models equally. The best fitting model included a factor of block and cue predictiveness, $BF_{10} = 1.65e + 94$.

Compared to the best fitting model, the closest fitting model with an effect of condition was the model with factors block, cue predictiveness, and condition. $BF_{10} = 2.88e + 93$. Therefore, the analysis suggests that a model that suggested no difference between the conditions on attention paid to cues is approximately 5.73 times more likely than a model assuming a difference.

Discussion

Experiment 3 failed to fully replicate the protection from uncertainty effect in responding in the *gradual change* condition, however the effect *was* observed in the *mixed change* condition. The reason for this discrepancy between the two experiments, and why the effect would still appear in the *mixed change* condition but not the *gradual change* condition is not immediately clear. It is possible the effect is simply stronger in the *mixed change* condition, though there was no evidence of a difference between the *mixed change* and *gradual change* conditions in stage two (despite there being a difference between the *mixed change* and *sudden change* condition).

Given that the protection from uncertainty effect was originally observed in the *gradual change* condition, and replicated (in a sense) in the *mixed change* condition, it seems likely that the failure to detect a protection from uncertainty effect in the *gradual change* condition may reflect a Type II error. It is clear though that further experimentation on the efficacy of the *gradual change* condition to produce the protection from uncertainty effect is warranted.

More importantly, these findings suggest that the gradual move to uncertainty is not necessary to achieve protection from uncertainty. These results partially align with the findings of Gureckis and Love (2009). In their task, participants completed a reinforcement learning task with either a low, moderate, or high level of reward variability. They found that, when participants experienced a moderate level of uncertainty, they appeared to be more diligent in solving the task, and thus ended up with better overall rewards than participants with either high or low uncertainty. The current task mirrors this finding to a degree. When participants experienced moderate

uncertainty, they made more high-value responses under a following period of high uncertainty than participants who had experienced no uncertainty.

The more surprising result is that the pattern of participants' attention did not match their behaviour. Previous studies have shown that attention and response behaviour are closely linked (Krajbich & Rangel, 2011). However, the results of Experiment 3 suggest that they may be separable in specific instances, or that attention may be indexing the EE trade-off in a different way than originally thought. Experiment 4 attempted to replicate this difference in responding and attentional behaviour.

Experiment 4

Experiment 4 aimed to focus on establishing the existence (or lack thereof) of the protection from uncertainty effect in both responding and attention[7]. The *gradual change* condition from Experiments 2 and 3 was dropped from Experiment 4, with the focus only on the *sudden change* and *mixed change* conditions. This provided two benefits: first, the *sudden change* and *mixed change* conditions would be directly compared in the omnibus analysis (minimising the need to run follow-up analyses). Second, it allowed for an increase in the number of participants in each condition, while also reducing the time needed to gather data.

One issue noticed after running Experiment 3 was the method by which participants were excluded from analysis, with participants excluded based on their overall performance in stage one. However, it is likely that it was easier for participants to learn the relationship between cues and high-value responses in the *sudden change* condition compared to the other two conditions, as rewards in the *sudden change* did not vary in stage one. Given this, the exclusion criterion may have been biased to exclude fewer poor performers in the *sudden change* condition. To address this, in Experiment 4 the initial period of stage one (where both conditions experience no uncertainty) was extended to three blocks, and participants were excluded based *only*

[7] The preregistration for this experiment can be found at https://osf.io/fvuw3/. Again, the analysis was changed to regression as in Experiment 3, and the results of the original analysis can be seen in the OSF repository

on their performance in these initial three blocks. To keep the length of the experiment within an hour, the length of stage two was shortened by two blocks, from six blocks to four.

The eye-gaze exclusion criteria were modified for Experiment 4. In Experiment 3, participants were excluded based over the whole experiment. However, it is possible that participants may have had greater or fewer tracking trials in each section of the task, depending on their level of motivation in each section. To avoid this issue, participants were excluded if they had fewer than 50% trials with a fixation on either cue in the three sections of the task: blocks 1 to 3 (where there is no uncertainty in either condition), blocks 4 to 9 (where only the *mixed change* condition experiences uncertainty), and stage two (where both conditions experience uncertainty).[8]

Based on the results of Experiment 3, it was predicted that participants in the *mixed change* condition would make fewer low-value responses than participants in the *sudden change* condition when moved to a high level of uncertainty in stage two. Similarly, it was also predicted that there would not be any difference between conditions in their proportion of trial time looking at cues in stage two, indicating a disconnect between responding and attention.

Method

Participants. One hundred and twenty one students from UNSW Sydney participated for course credit. The mean age was 19.7 years (SD = 3.09); 63 participants identified as female and 57 as male[9]. Exclusion criteria were preregistered to exclude participants with an average of under 55% high-value responding during block 1 to 3. Twenty-four participants were excluded due to this criterion. A further five participants were excluded due to having fewer than 50% of trials with at least one fixation on a cue in one or more of the three specified sections. Testing continued until

[8] While a similar analysis could be done on the data in Experiment 3, this would be in opposition to our preregistered exclusion criteria.

[9] Due to experimenter error, the data from one participant was excluded from the *mixed change* condition for failing to meet the eye-tracking criteria was lost, and their data were not included for calculation of mean age or standard deviation.

there were 45 participants in each group that did not have to be excluded, leaving 90 participants for the final analysis. The two highest performing participants were paid $20 after data collection had finished.

Materials. All materials were the same as those used in Experiments 1 and 3.

Design. The design of Experiment 4 was identical to that of Experiment 3, with the exception that participants were only assigned to two conditions: the *sudden change* and *mixed change* conditions.

Procedure. The procedure of Experiment 4 was identical to that of Experiment 3, with the exception that stage one now ran for nine blocks of 32 trials each, with the first three blocks (rather than the first two) of the experiment identical in both conditions, in that there was no variability in the rewards during this period. This was to ensure that participants had ample opportunity to learn the relationship between cues and responses in both conditions in stage one, with the aim to reduce the amount of participants that needed to be excluded from the experiment. At the start of block 4, uncertainty was introduced for the *mixed change* condition in the same fashion as in Experiment 3. To keep the task within an hour, stage two now ran for only four blocks.

Results

Data were split into blocks of 32 trials for analysis. Trials that were longer than two standard deviations above or below the mean trial time were excluded from analysis. When this criterion was applied, a median of 2 trials and mean of 1.7 trials per block were removed. The key behavioural results are summarised for brevity in this section, but the full results of all tests can be seen in Supplementary E.

Response behaviour. Response data from Experiment 4 can be seen in Figure 7. As in Experiment 3, only blocks 4 to 9 of stage one were analysed, excluding blocks 1 to 3 where both conditions experienced no reward variability[10].

---

[10] Cursory analysis of blocks 1 to 3 revealed no significant effect of condition, $(\chi^2(2) = 1.55, p = .214$, suggesting that the three conditions were fairly well matched on performance during blocks 1 to 3.

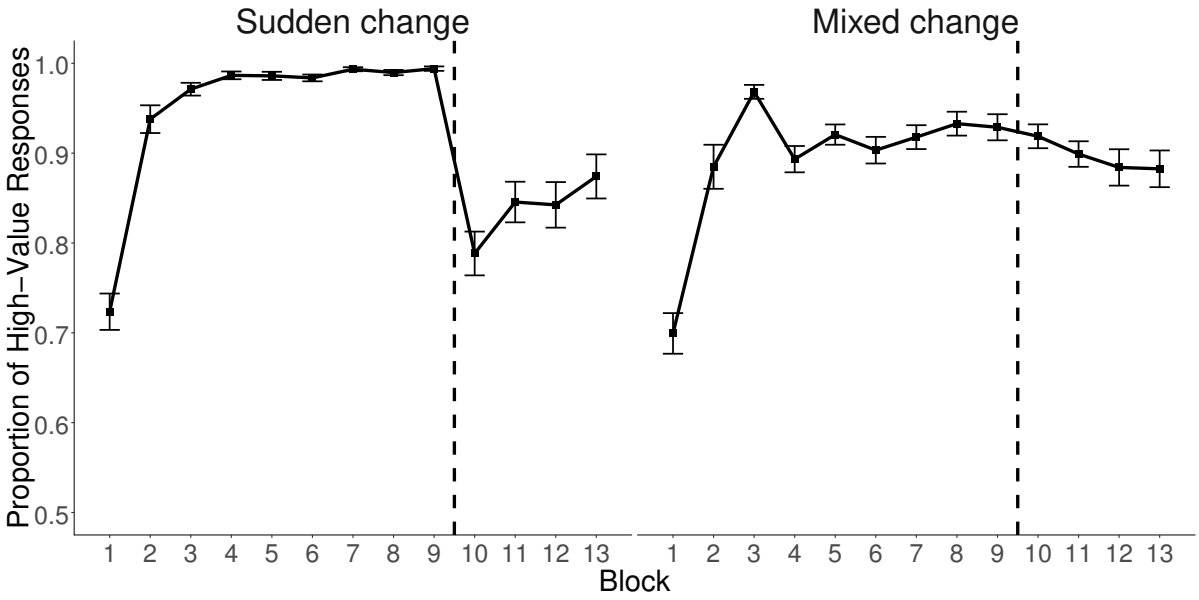

Figure 7. The choice data from both conditions of Experiment 4. Stage one occurred during blocks 1 through 9, and stage two occurred on blocks 10 through 13, with the dotted vertical line indicating where the stages changed. In the *sudden change* condition, rewards became more variable in stage two. In the *mixed change* condition, rewards became moderately variable in block four of stage one, then more variable in stage two. Error bars represent standard error of the mean.

As expected, both conditions showed a clear improvement in the proportion of high-value responses made over the course of stage one, $\chi^2(5) = 17.92, p = .003$, with participants in the *sudden change* condition making more high-value responses overall compared to participants in the *mixed change* condition, $\chi^2(1) = 101.02, p < .001$. In the transition between stages one and two, the findings from previous experiments were replicated, with participants in the *sudden change* condition making more high-value responses in block 9 compared to participants in the *mixed change* condition, $b = 2.62, z = 5.63, p < .001$, 95% PLCI $[1.77, 3.63]$, but fewer in block 10, $(b = -1.29, z = -5.21, p < .001$, 95% PLCI $[-1.80, -0.81]$.

There was an overall main effect of condition in stage two, with participants in the *mixed change* condition making significantly more high-value responses than participants in the *sudden change* condition, $\chi^2(1) = 5.73, p = .017$. However, unlike in Experiments 1, 2, and 3 participants in the *sudden change* condition managed to improve substantially after the initial onset of uncertainty at the beginning of stage one. Though participants in the *sudden change* condition still experienced the marked

decrease in high-value responding in the first block of stage two, $\chi^2(3) = 47.38, p < .001$, they quickly returned to making high-value responses at around the same level as participants in the *mixed change* condition by block 13, $b = 0.01, z = 0.28, p = .782$, 95% PLCI $[-0.61, 0.83]$. In other words, participants in the *sudden change* condition appeared to explore the low-value response immediately after the onset of uncertainty, but quickly began returned to exploiting the high-value response after this initial exploration.

Attention. Eye-gaze data were processed in the same manner as in Experiment 1. Eye-gaze data for Experiment 4 can be seen in Figure 8. As with the response data, only blocks 4 to 9 of stage one were analysed[11].

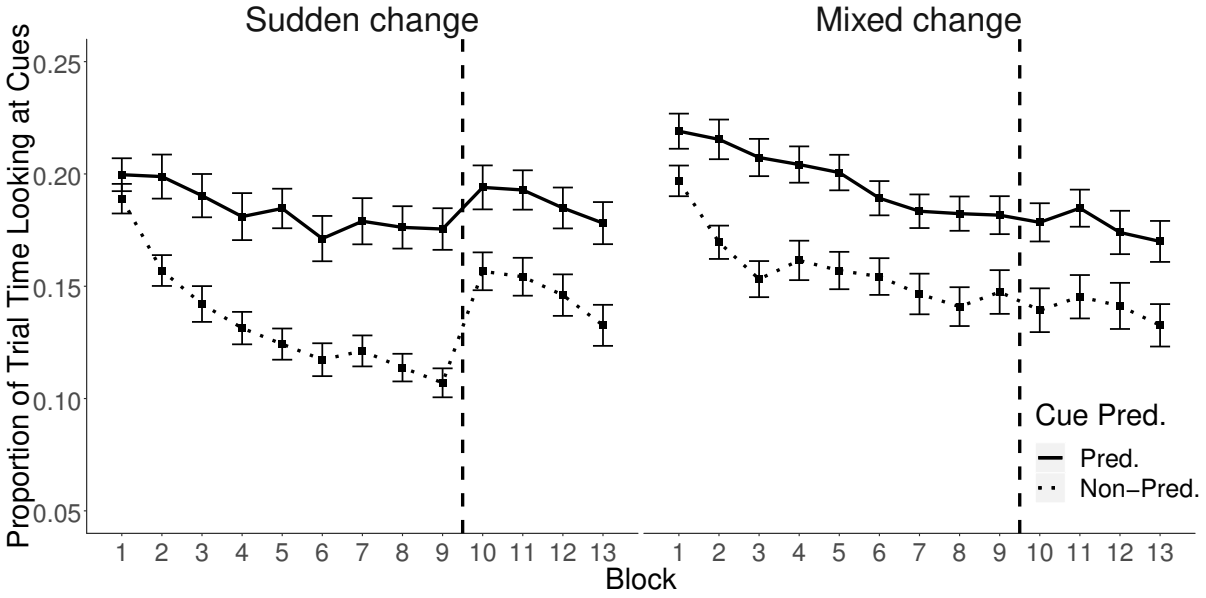

Figure 8. The eye-gaze data from both conditions in Experiment 4. Stage one occurred during blocks 1 through 9, and stage two occurred on blocks 10 through 13, with the dotted vertical line indicating where the stages changed. In the *sudden change* condition, rewards became more variable in stage two. In the *mixed change* condition, rewards became moderately variable in block four of stage one, then more variable in stage two. Error bars represent standard error of the mean.

[11] Cursory analysis of blocks 1 to 3 revealed no significant differences between the conditions in the proportion of trial time looking at cues, $\chi^2(1) = 2.95, p = .086$.

In stage one, participants in the *mixed change* condition spent a greater proportion of the trial attending time to cues than participants in the *sudden change* condition, $\chi^2(1) = 6.09, p = .014$, indicating greater exploration of the cues. Similarly, participants in the *sudden change* condition were better able to bias their attention to the predictive cue over the non-predictive cue compared to participants in the *mixed change* condition, $\chi^2(1) = 18.21, p < .001$. That is, participants in the *mixed change* condition showed a greater proportion of looking time to the non-predictive cue than participants in the *sudden change* condition, $b = 0.02, t(88) = 3.24, p = .002$, 95% PLCI $[6.36e-3, 2.58e-2]$, but there was no significant difference between the conditions in the proportion of trial time looking at the predictive cue, $b = 6.163e-3, t(88) = 1.09, p = .278$, 95% PLCI $[-4.89e-3, 1.72e-2]$. Overall, these results indicated greater attentional exploitation in the *sudden change* condition in stage one.

The transition between stages showed that participants in the *sudden change* condition increased their proportion of trial time looking at the non-predictive cue more than participants in the *mixed change* condition from block 9 to block 10, $\chi^2(1) = 4.86, p = .028$. This suggested that participants in the *sudden change* condition re-evaluated the usefulness of the non-predictive cue under uncertainty.

Again, there was no significant difference in attention between conditions in stage two, $\chi^2(1) = 0.73, p = .394$, with participants appearing to spend the same proportion of the trial looking at cues in both conditions. To assess the evidence against a difference between conditions in stage two, a Bayesian ANOVA was run, with fixed factors of condition, block, and cue predictiveness, and a random factor of participant. The best fitting model included a factor of block and cue predictiveness, $BF_{10} = 2.39 + e37$. Compared to the best fitting model, the closest fitting model with an effect of condition was the model with factors block, cue predictiveness, and condition, $BF_{10} = 9.60 + e36$. Therefore, the analysis suggests that a model that suggested no difference between the conditions on attention paid to cues is approximately 2.49 times more likely than a model assuming a difference.

Discussion

Experiment 4 aimed to confirm the pattern of response behaviour and attention seen in Experiment 3. Experiment 4 had two conditions, the *sudden change* and *mixed change* condition, and employed stricter and less biased exclusion criteria. Broadly, this aim was met: participants who had experienced a moderate level of uncertainty in the *mixed change* condition made more high-value responses than participants in the *sudden change* condition when exposed to a high level of uncertainty in stage two. Furthermore, there was no evidence to suggest that there was a difference between conditions in the proportion of trial time looking at cues in stage two, despite the difference in the level of high-value responses.

Unexpectedly, participants in the *sudden change* condition returned to making the high-value response at the same level as participants in the *mixed change* condition in stage two, contrary to the results of previous experiments. This may have occurred due to Experiment 4's stricter exclusion criterion for responding. As it was ensured that participants must have over 55% high-value responding over blocks 1 to 3, only participants who were able to quickly pick up the relationship between cues and responses were included for analysis. Therefore, though participants in the *sudden change* condition were still affected by the sudden onset of uncertainty in stage two, these high performing participants may have been able to quickly learn that the relationships between cues and responses from stage one were still the same in stage two.

In terms of the attentional data, there was evidence in favour of an attentional EE trade-off during stage one. That is, participants in the *sudden change* condition showed a greater bias in attention between the predictive cue and the non-predictive cue compared to that shown for participants in the *mixed change* condition, while participants in the *mixed change* condition paid more attention to cues overall. However, there was no difference between conditions in stage two, with participants in the two conditions paying roughly equal attention to the cues. Importantly, the change in attention at the onset of stage two in the *sudden change* condition observed in Experiment 3 was supported. That is, following the introduction

of uncertainty, participants in the *sudden change* condition showed the largest increase in attending to the non-predictive cue. This indicated an increase in exploratory attention at the onset of uncertainty, and may reflect a re-evaluation of information previously thought to be uninformative.

General discussion

The current paper explored the impact of expected and unexpected uncertainty on the exploitation/exploration trade-off (EE trade-off). Four conditions were employed across four experiments that differed in the way uncertainty was introduced to participants, and the impact of this uncertainty on responding and attention was assessed. In each experiment, participants were shown two cues and asked to choose between two possible responses. One of those cues predicted which response would on average give more points, while the other was non-informative. The way that participants were exposed to uncertainty was manipulated in an initial stage (stage one), before all participants were exposed to a high level of uncertainty in a following stage (stage two). Crucially, the relationship between cues and responses was identical in both stages. Therefore, any sustained attempt to explore once the participant had learned the relationship between cues and responses had been learned was counter to the goal of maximising rewards.

It was found that when a high level of uncertainty was introduced unexpectedly in stage two, exploration increased dramatically. In contrast, when participants had experienced a continued period of uncertainty in the first stage, their pattern of responding was more exploitative in stage two. That is, the extent to which participants showed exploitative response behaviour (favouring the high-value response) was modulated by the type of uncertainty they had experienced. With a high level of uncertainty throughout the first stage (i.e., in the *always uncertain* condition; Experiments 1 and 2), participants tended to fail to learn the relationship between cues and high-value responses, responding at a sub-optimal rate throughout the task. However, when the level of uncertainty was more moderate, such that participants could

learn the association between cues and high-value responses (e.g., in the *gradual uncertainty* condition; Experiments 2 and 3) the likelihood of seeing more high-value responding increased in stage two. We term this a *protection from uncertainty effect*, where a moderate level of uncertainty actually benefits high-value responding in uncertain conditions, compared to when uncertainty is experienced more suddenly. This finding aligns well with Cohen et al. (2007), who postulated that participants should increase their exploration following a sudden change in the environment, as it may indicate that there is new information to learn in the environment. By contrast, when participants are gradually exposed to a high level of uncertainty, they should be more willing to slowly update their understanding of the environment, accounting for uncertainty while broadly maintaining the same response strategy (Courville et al., 2006).

Relation to the partial reinforcement extinction effect

The protection from uncertainty effect may be considered a related phenomenon of a well established associative effect known as the *partial reinforcement extinction effect* (Sheffield, 1949; Weinstock, 1954). The partial reinforcement extinction effect describes a somewhat paradoxical finding that participants who are trained on an inconsistent cue/outcome pairing are more resistant to later extinction (unlearning of that cue/outcome association) than participants who are trained on a consistent cue/outcome pairing. Similar to the protection from uncertainty effect, the partial reinforcement extinction effect suggests that unexpected change (i.e., when a fully reinforced cue/outcome contingency is suddenly degraded) can alter behaviour.

The idea that the partial reinforcement extinction effect may be explained as a consequence of unexpected change has been explored in the literature (Blanco & Moris, 2017; Gershman et al., 2013; Haselgrove, Aydin, & Pearce, 2004; Pearce, Redhead, & Aydin, 1997). Called the *trial-based* account of the partial reinforcement extinction effect (Harris & Bouton, 2020), it argues that a shift in context between learning and extinction is critical to facilitate the effect. Once a context shift is detected

(e.g., when a fully reinforced cue/outcome association begins to be extinguished), learners become primed to begin learning about new associations in the environment, so behaviour changes quickly. If no context shift is detected (as might be the case for a partially reinforced cue/outcome association moving to extinction), the rate of learning for new associations remains low, so behaviour changes slowly.

This account dovetails nicely with the idea that unexpected uncertainty induces exploration: when there is a sudden change in the reinforcement ratio from learning to extinction, a context shift is detected, and exploration occurs. However, a more popular competitor to explain the partial reinforcement extinction effect is the *time-accumulation* account. In the time-accumulation account, the partial reinforcement extinction effect is thought to occur as a consequence of increased time between outcomes in training (Gallistel & Gibbon, 2000). For example, if an animal is rewarded every time it sees a cue, and that cue is presented every 10 seconds, then the animal also learns that it can expect a reward every 10 seconds. If the animal is only rewarded on 50% of trials, then it comes to expect the reward only every 20 seconds. Therefore, when the experimenter tries to extinguish the cue/outcome association with the animal, it takes approximately twice as long for the animal to realise that no reward is coming if it had received partial reinforcement of cues and outcomes compared to if it had received full reinforcement. The time-accumulation account has been so successful in explaining the partial reinforcement extinction effect that even proponents of trial-based accounts have acknowledged that context change alone without the effect of time is insufficient to fully explain the effect (Gershman et al., 2010).

However, in the case of the protection from uncertainty effect it appears unlikely (though not impossible) that time between rewards had an effect on increasing exploratory behaviour in stage two. As participants were rewarded on every trial with only the magnitude of reward changing, under the assumption that trials took an approximately equal amount of time for each participant there should be no difference in the interval between rewards in any of the conditions. Despite this, those in the *sudden change* condition still showed a greater increase in exploratory responding than

participants in the *gradual change* and *mixed change* conditions. Instead, it seems more likely that the context established by the noisy rewards in stage one for the *gradual change* and *mixed change* condition protected participants from inferring any context shift in stage two (as they expected the presence of uncertainty). Indeed, a shift in context may be one of the reasons that unexpected uncertainty motivates exploration: once an organism detects that the environment has changed (i.e., the context has shifted), they should explore for new relationships between the cue and the outcome by increasing their attention to cues (Easdale et al., 2019), and increasing their rate of learning (Blanco & Moris, 2017).

It should be noted that a key difference between the two effects is that in the partial reinforcement extinction effect, it is beneficial to explore following unexpected uncertainty, as the relationship between the cue and the outcome does actually change. By contrast, in the protection from uncertainty effect, it is harmful to explore following unexpected change, as the value of making the high-value response does not change. This distinction is important, as the two tasks form complementary bodies of evidence for the impact of sudden change on behaviour. In the partial reinforcement extinction effect, those who do not interpret a change in the task are disadvantaged, as the behaviour they should be performing changes. On the other hand, in the protection from uncertainty effect, those who *do* experience the change are disadvantaged, as they are sent on a proverbial "wild goose chase" for new response strategies despite theoretically already knowing the strategy to maximise rewards. The combination of these two tasks shows that both expected and unexpected uncertainty can produce patterns of behaviour that is directly counter to an agent's goal of performing appropriately in the task (by either under-motivating or over-motivating exploration). Taken together, the partial reinforcement extinction effect and the protection from uncertainty effect provide strong support for the idea that unexpected uncertainty may promote exploratory behaviour.

The Attentional E/E trade-off

In terms of the attentional data, there was clear evidence for an attentional EE trade-off in stage one. Participants in the *sudden change* condition generally showed greater prioritisation of predictive information over non-predictive information compared to participants in conditions with greater uncertainty, demonstrating attentional exploitation. In Experiment 4, evidence was also found for attentional exploration, with participants in the *mixed change* condition showing greater attention to cues overall in stage one compared to participants in the *sudden change* condition. These findings align with previous work on the attentional EE trade-off (Beesley et al., 2015; Easdale et al., 2019; Walker et al., 2017).

It was also expected that in stage two, the EE trade-off in the attentional data would match the EE trade-off in the response data. That is, it was expected that when participants demonstrated exploratory responding, they would also demonstrate an exploratory pattern of attention (and similarly for exploitation). To some extent, this was true: participants in the *sudden change* condition showed a significant increase in attending to cues as well as a significant decrease in high-value responding at the onset of uncertainty. However, between conditions there was no evidence of a difference in attending to cues in stage two, despite a difference in response behaviour. Indeed, there was moderate evidence *against* a difference between conditions in attention.

This finding is at odds with the majority of previous research on the interaction between attention and choice. A number of papers have argued that attention is closely tied to choice behaviour, such that one can predict choices via attention (e.g Krajbich & Rangel, 2011; Stewart, Gächter, Noguchi, & Mullett, 2016; Stewart, Hermens, & Matthews, 2016). Due to this, the current results should be interpreted with some caution. Though our analysis used Bayesian methods to provide evidence for the null, the evidence is at best only moderate.

Keeping in mind the above caveat, one potential reason that this pattern of behaviour occurred is that it may be less costly for participants to explore with their attention than with their responses. While trying the low-value response often risks an

immediate loss of points that directly harms the chance of receiving a monetary reward at the end, taking the effort to attend to both cues on each trial ensures that any patterns that emerge in the cue/response associations are not missed. Therefore, participants in the *gradual change* and *mixed change* conditions may have continued to explore with their attention in stage two, matching the participants in the *sudden change* condition, even though they were able to perform the task reasonably well.

One issue with this explanation is that attention appeared to diverge between conditions in stage two of Experiment 1, with participants in the *sudden change* condition showing greater attentional exploitation than participants in the *always uncertain* condition (a finding also seen in Easdale et al., 2019). This result is difficult to interpret in the face of the other two results (where response behaviour, not attention, diverged between conditions). Given that no participants were excluded based on response behaviour in Experiment 1, it is possible that more participants failed to learn any relationship between cues and responses at all in the *always uncertain* condition compared to the *sudden change* condition. If this were the case, the bias towards attending to the predictive cue over the non-predictive cue may have been attenuated in the *always uncertain* condition in Experiment 1.

Before concluding the discussion on attention, it should be noted the definition of an attentional EE trade-off used in this paper may be considered at odds with the traditional definition of the EE trade-off. In the choice domain, if one wishes to explore, it is generally necessary to forego exploiting (and vice-versa). That is, a participant that has a single selection to allocate between multiple arms cannot choose to explore an unknown arm while also exploiting an arm for reward. By contrast, this strict dichotomy does not exist in our operationalisation of an attentional EE trade-off. We have defined exploitative attention as attending preferentially to predictive cues, and exploratory attention as increasing attention to cues overall. Under this definition, it is possible for participants both to exploit with their attention (preferentially attending to predictive cues), while also exploring (increasing attention to all cues). Indeed, this is the pattern seen in the data: when participants explore, they increase

the eye-gaze on cues overall, but not an the expense of a selective bias to predictive over non-predictive cues (see also Beesley et al., 2015; Easdale et al., 2019). The question is raised then that if a participant can both explore and exploit simultaneously with their attention, does an EE trade-off exist at all?

While ostensibly a fair criticism, it relies on two fundamental premises: that the EE trade-off exists only at the level of obtained outcomes; and that exploration and exploitation are entirely separate processes. However, a recent sweeping review on the EE trade-off literature by Mehlhorn et al. (2015) argued that neither of these premises are true. Mehlhorn et al. (2015) postulated that the EE trade-off exists across three dimensions: obtained outcomes (either exploiting for rewards or exploring for information); behaviour of the agent (exploiting by continuously selecting a single arm or exploring by spreading selections over multiple arms); and the value and uncertainty related to choice options (exploiting by selecting arms with high subjective values with low uncertainty or exploring by selecting arms with low subjective values and high uncertainty). Furthermore, Mehlhorn et al. (2015) postulated that these processes are not dichotomous, but exist on a continuum. That is, rather than behaviour being definitively categorised as either exploitative or exploratory, behaviour moves along this continuum, shifting from more exploratory behaviour to more exploitative, and vice versa.

The attentional EE trade-off becomes clearer when considered through Mehlhorn et al.'s (2015) lens. If a participant shows a attentional bias towards predictive cues, and totally ignores other cues, they show a clear preference for attending to high value (in this case, informative), low uncertainty cues, and this would indicate "strong exploitation". By contrast, if a participant shows an attentional bias towards predictive cues, but also dedicates some attention to other cues with low value or high uncertainty, the participant shifts towards the exploratory end of the spectrum in a style of "weak exploitation" or "weak exploration". Under this more modern view of the EE trade-off, it is clear that an EE trade-off can exist in attention as we have defined it here.

Defining unexpected uncertainty and a continuum of unexpectedness

In the current paper, unexpected uncertainty was operationalised as the unannounced introduction of noise in rewards at a set point in the task. By contrast, expected uncertainty was operationalised either as the presence of noise in the rewards that gradually increased in magnitude, or as the presence of noise distributed with a central peak (i.e., the *gradual change* and *mixed change* conditions respectively). From these conditions, it may seem reasonable to define unexpected uncertainty as an increase in environmental uncertainty that occurs without warning. By contrast, expected uncertainty is environmental uncertainty that occurs at a level known or expected by the participant. Following this, it appears sensible to consider these two states to be mutually exclusive (i.e., uncertainty is either expected or it is not).

However, it may be more appropriate to conceptualise expected and unexpected uncertainty as two extremes of a continuum, rather than as two dichotomous states. Indeed, such an interpretation provides a sensible lens through which to view the current data. For example, consider the *sudden change* and *gradual change* conditions. Though participants in both the *sudden change* and *gradual change* conditions experienced some form of unexpected uncertainty, the extent to which that uncertainty was unexpected may have differed between these two conditions, potentially explaining why participants in the *gradual change* condition were protected from the need to explore in stage two. That is, in the *gradual change* condition, each time the range of outcomes increased in stage one, the participant should have experienced unexpected uncertainty (as new rewards were presented that had not been seen before). However, it is clear that these small increases in uncertainty over the course of stage one motivate exploration far less than the single dramatic increase in uncertainty present in the *sudden change* condition. The implication of this is that there may be *continuum of unexpectedness* (from expected to unexpected), with greater unexpectedness leading to greater exploratory behaviour.

This conceptualisation of a continuum of unexpectedness aligns well with the original work of Yu and Dayan (2005). Yu and Dayan (2005) originally defined

unexpected uncertainty as the occurrence of an outcome outside the expected range of outcomes. For example, if one expects rewards in the range of 14-16, and they receive a 10, this unexpected score creates unexpected uncertainty. They also argued that the extent to which a reward is unexpected (i.e., the further away a reward is from the expected range of rewards) will impact the level of exploration. For example, a score of 5 when one is expecting 14-16 would induce more unexpected uncertainty than a score of 12. Under this definition, it is clear that those in the *sudden change* condition would be more likely to explore in stage two than those in the *gradual change* condition as there is a much greater subversion of expectations about reward values in the *sudden change* condition. Indeed, this is what we observed in the protection from uncertainty effect, with those who experienced small, consistent changes in the reward distribution exploring far less at the start of stage two than those who experienced one significant change, despite experiencing the same absolute level of uncertainty. The implication of this finding is that it suggests that experience of unexpected uncertainty may differ based on the magnitude of deviance from expected rewards, suggesting that these two states may lie at two ends of a continuum of unexpectedness.

This idea of a continuum of unexpectedness dovetails nicely with previous work on the EE trade-off by Mehlhorn et al. (2015). As previously mentioned in the discussion on the attenional EE trade-off, Mehlhorn et al. (2015) have argued that exploration and exploitation themselves may not be dichotomous processes, but two ends of one continuum. Similarly, the differences in exploratory behaviour shown when participants experience expected versus unexpected uncertainty may represent one part of a larger continuum between exploration and exploitation, with unexpected uncertainty lying closer towards exploration, and expected uncertainty lying closer towards exploitation.

This idea of a continuum of unexpectedness opens up some interesting possibilities for future research. For example, computational models that try to capture the impact of unexpected uncertainty on the EE trade-off may have to consider whether they will capture unexpected uncertainty in a discrete fashion (i.e., either an outcome is

unexpected or it is not), or in a linear fashion (i.e., on a continuum from unexpected to expected uncertainty). Similarly, future experiments may consider trying to more explicitly manipulate the extent to which uncertainty is unexpected to see how exploratory behaviour changes. For example, an experimenter might warn the participant that uncertainty will occur at some point in the task (reducing the level of unexpectedness) to see how that changes exploration.

Differences in exclusion criteria

It is worth briefly addressing why the exclusion criteria employed across the current study have changed across each experiment. The goal of the exclusion criteria was to avoid analysing participants who did not learn about the relationship between cues and responses in the task, or (in the case of eye-tracking) had too few data points recorded to be meaningful. The evolution of exclusion criteria reflect the natural process of trying to achieve this goal across multiple experiments. We have striven to make these exclusion criteria and the process that lead up to their inclusion entirely transparent (including preregistering the criteria for Experiments 3 and 4). Ultimately, it has been argued that the criteria employed are fit for the goals of the study, but importantly the reader is also given adequate information to decide for themselves whether they believe the criteria are appropriate.

Conclusions

It has been shown across four experiments how the sudden onset of uncertainty can motivate exploratory behaviour in response behaviour and attention. When participants were suddenly exposed to uncertainty, they showed a dramatic increase in exploring different responses and different cues. Furthermore, it was found that when uncertainty is expected to be present and the best response strategy had been learned, participants appeared to be protected from the effects of uncertainty on exploration in responding, but not in attention. These findings suggest that how uncertainty is introduced to decision-makers influences how they perform the EE trade-off, extending earlier work by Cohen et al. (2007) and Yu and Dayan (2005), and

that attention and response behaviour may index different aspects of the EE trade-off. This conclusion has implications for how the EE trade-off should be conceptualised, providing support to the notion that the EE trade-off may represent a continuum of behaviour, rather than a dichotomy.

Acknowledgements

This work was supported by an Australian Research Council Discovery Project (DP140103268), and a Research Training Program scholarship from the Australian Department of Education and Training. The authors would also like to thank Lachlan Kay and Nicole Baz for their help with data collection.

Reference list

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

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.

Blanco, F., & Moris, J. (2017). Bayesian methods for addressing long-standing problems in associative learning: The case of PREE. *The Quarterly Journal of Experimental Psychology*, 1–52.

Bradt, R. N., Johnson, S., & Karlin, S. (1956). On sequential designs for maximizing the sum of n observations. *The Annals of Mathematical Statistics*, *27*(4), 1060–1074.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

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.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*(7), 294–300.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879.

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

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (Second ed.). Thousand Oaks, California: Sage.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological review*, *107*(2), 289.

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, Learning, and Extinction. *Psychological Review*, *117*(1), 197–209.

Gershman, S. J., Jones, C. E., Norman, K. A., Monfils, M.-H., & Niv, Y. (2013). Gradual extinction prevents the return of fear: implications for the discovery of state. *Frontiers in Behavioral Neuroscience*, *7*, 1–6.

Gershman, S. J., & Niv, Y. (2012). Exploring a latent cause theory of classical conditioning. *Learning & Behavior*, *40*(3), 255–268.

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Statistics*, *41*(2), 148–177.

Gureckis, T. M., & Love, B. C. (2009). Learning in noise: Dynamic decision-making in a variable environment. *Journal of Mathematical Psychology*, *53*(3), 180–193.

Harris, J. A., & Bouton, M. E. (2020). Pavlovian conditioning under partial reinforcement: The effects of nonreinforced trials versus cumulative conditioned stimulus duration. *Journal of Experimental Psychology: Animal Learning and Cognition*.

Haselgrove, M., Aydin, A., & Pearce, J. M. (2004). A partial reinforcement extinction effect despite equal rates of reinforcement during pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*(3), 240.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., & Others. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14).

Konstantinidis, E., Taylor, R. T., & Newell, B. R. (2018). Magnitude and incentives: revisiting the overweighting of extreme events in risky decisions from experience. *Psychonomic Bulletin & Review*, *25*, 1925–1933.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852–13857.

Le Pelley, M. E., Beesley, T., & Griffiths, O. (2011). Overt attention and predictiveness in human contingency learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*(2), 220–229.

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

Manohar, S. G., & Husain, M. (2013). Attention as foraging for information and value. *Frontiers in Human Neuroscience*, *7*, 711.

Marshall, L., Mathys, C., Ruge, D., de Berker, A. O., Dayan, P., Stephan, K. E., & Bestmann, S. (2016). Pharmacological fingerprints of contextual uncertainty. *PLoS Biology*, *14*(11), 1–31.

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., ... Gonzalez, C. (2015). Unpacking the exploration-exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*, 191–215.

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes Factors for Common Designs [Computer software manual].

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *Journal of Neuroscience*, *35*(21), 8145–8157.

Payzan-LeNestour, E., Dunne, S., Bossaerts, P., & O'Doherty, J. P. (2013). The neural representation of unexpected uncertainty during value-based decision making. *Neuron*, *79*(1), 191–201.

Pearce, J. M., Redhead, E. S., & Aydin, A. (1997). Partial reinforcement in appetitive pavlovian conditioning with rats. *The Quarterly Journal of Experimental Psychology: Section B*, *50*(4), 273–294.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.r-project.org/`

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

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Symposium on Eye Tracking Research & Applications*, 71–78.

Schulz, E., Konstantinidis, E., & Speekenbrink, M. (2018). Putting bandits into context: How function learning supports decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *44*(6), 927–943.

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*(3), 233–250.

Sheffield, V. F. (1949). Extinction as a function of partial reinforcement and distribution of practice. *Journal of Experimental Psychology*, *39*(4), 511–526.

Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science*, *7*(2), 351–367.

Stewart, N., Gächter, S., Noguchi, T., & Mullett, T. L. (2016). Eye movements in strategic choice. *Journal of Behavioral Decision Making*, *29*(2-3), 137–156.

Stewart, N., Hermens, F., & Matthews, W. J. (2016). Eye movements in risky choice. *Journal of Behavioral Decision Making*, *29*(2-3), 116–136.

Walker, A. R., Le Pelley, M. E., & Beesley, T. (2017). Exploitative and exploratory attention in a four-armed bandit task. *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, 3484–3489.

Walker, A. R., Luque, D., Le Pelley, M. E., & Beesley, T. (2019). The role of uncertainty in attentional and choice exploration. *Psychonomic Bulletin & Review*. (Advanced online publication)

Weinstock, S. (1954). Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *Journal of Comparative and Physiological Psychology*, *47*(4), 318–322.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692.