The role of uncertainty in attentional and choice exploration

Adrian R. Walker[1], David Luque[2,1], Mike E. Le Pelley[1], & Tom Beesley[3,1]

1. **UNSW Sydney, Australia**
2. **Universidad Autónoma de Madrid, Spain**
3. **Lancaster University, UK**

Address correspondence to:

Adrian R. Walker

School of Psychology, UNSW Sydney, NSW 2052, Australia.

adrian.walker@unsw.edu.au

Abstract

The exploitation-exploration (EE) trade-off describes how, when making a decision, an organism must often choose between a safe alternative with a known pay-off, and one or more riskier alternatives with uncertain pay-offs. Recently, the concept of the EE trade-off has been extended to the examination of how organisms distribute limited attentional resources between several stimuli. This work suggests that when the rules governing the environment are certain, participants learn to "exploit" by attending preferentially to cues that provide the most information about upcoming events. However, when the rules are uncertain, people "explore" by increasing their attention to all cues that may provide information to help in predicting upcoming events. In the current study, we examine how uncertainty affects the EE trade-off in attention using a contextual two-armed bandit task, where participants explore with both their attention and their choice behavior. We find evidence for an influence of uncertainty on the EE trade-off in both choice and attention. These findings provide support to the idea of an EE trade-off in attention, and that uncertainty is a primary motivator for exploration in both choice and attentional allocation.

Keywords: Attention, Human Associative Learning, Exploration, Uncertainty

When making a decision about the behavior to perform in a given situation, an agent can choose to use its current knowledge to maximize short-term gains, or search for new information to maximize long-term gains (Cohen, McClure, & Yu, 2007). Imagine a person eating at their favorite restaurant, where they can choose to either order a meal they have had before (exploiting their knowledge about existing choice values), or try something on the menu they have never eaten before (exploring the value of a new choice). Such *exploitation-exploration* (EE) trade-offs occur frequently in daily life, such as when deciding the route to travel in a car, or making financial investments (Mehlhorn et al., 2015).

Until recently, research on EE trade-offs had been restricted to the domain of choice behavior. As in the restaurant example above, this involves studying how decision makers allocate their choices between several alternatives. These studies have shown that uncertainty is key to motivating exploration of choice responses (Knox et al., 2012; Speekenbrink & Konstantinidis, 2015). Beesley, Nguyen, Pearson, and Le Pelley (2015) recently showed that uncertainty can also affect attentional processing in a manner consistent with the EE trade-off. In a modified version of the 'learned predictiveness design' (Le Pelley & McLaren, 2003), on each trial participants were presented two of four possible cues (cues A, B, X, & Y) and were required to select one of two possible responses (response 1 or response 2). Participants were then told whether their response was correct or incorrect. On each trial, one of the cues (A or B) predicted the correct response to make (the "predictive cue"), and the other cue (X or Y) was irrelevant (the "non-predictive cue"), providing no information on which response was correct. Critically, the certainty with which predictive cues indicated the correct response varied across groups of participants. For participants in the "certain" condition, the relationship between the predictive cue and the accuracy of the response was perfect: whenever cue A was present, response 1 was the correct response, and whenever cue B was present, response 2 was the correct response. For participants in the "uncertain"

condition, the relationship between the predictive cue and the correct response was probabilistic. For example, if cue A was present, response 1 would be correct on 67% of trials, with response 2 correct on the remainder.[1]

Beesley et al. (2015) used eye-gaze measurements to assess how participants attended to the cues in the two different groups. They found participants in the certain group would on average spend a greater proportion of the trial time looking at the predictive cue compared to the non-predictive cue (see also Le Pelley, Beesley & Griffiths, 2011; Rehder & Hoffman, 2005). They argued this reflected an exploitation strategy: in a stable, certain environment, participants directed attention towards the best available cue to exploit their knowledge of the contingencies. It was also found that participants in the uncertain group spent a greater proportion of the trial time looking at cues overall than participants in the certain group. Beesley et al. argued that this reflected an exploration strategy: participants faced with an uncertain environment opted to spend more time attending to cues, presumably in an attempt to gain new information from the cues that could allow for more accurate response selection in the future.

Beesley et al. (2015) also tracked participants' choices during the task. While participants in the certain condition generally learned to choose the response that led to a correct outcome on nearly 100% of trials, participants in the uncertain condition typically engaged in a *probability matching* strategy, matching their responses to the probability of receiving a correct outcome for each response (Shanks, Tunney, & McCarthy, 2002). While this seems an intuitive strategy, it leads to a suboptimal level of performance; the optimal

[1] 'Uncertainty' here refers to the stochastic nature of observed outcomes, such that the outcome of an event is unknown beforehand (e.g., will a flip of this coin land on heads or tails?), rather than uncertainty about the *process* that generates outcomes, which can come to be known over time (e.g., an observer can learn that the probability of flipping heads is .5).

strategy under these conditions is to always select the response that is successful on the majority of trials (Shanks et al., 2002).

Beesley et al. (2015) suggested that probability matching under uncertain conditions may reflect exploratory behavior in choice. Linking this to their attentional findings, they proposed that under conditions of uncertainty, participants explored with both their responses (by probability matching) and their attention (by increasing gaze time to cues) simultaneously. While exploring with both attention and responses can eventually contribute to receiving better outcomes, this exploration reflects different aspects of behavior in the task. When participants explore by changing response strategies, we assume they are exploring in the sense of learning about the *value of the outcome for each choice*. In contrast, when participants explore by increasing their attention to cues, we assume they are doing so to encode more information about the stimuli, to assist in learning the *usefulness of each cue in predicting the outcome*.

One issue with this explanation is that participants in Beesley et al.'s (2015) study received feedback to indicate which response was correct on each trial, regardless of the response that was made. Therefore, while participants might have been motivated to explore the cues for new information under uncertain conditions, there was no logical reason to try different responses in the task. This is clearly at odds with the standard EE trade-off problem: since participants received information about the value of both responses (after trying only one), they could explore (gain new information about the different choices) and exploit (pick the best-known choice) simultaneously, removing any trade-off between the two behaviors.

Another important facet of Beesley et al.'s (2015) procedure was that in their task, the difference in the long-run reward value of the optimal and suboptimal response was substantially smaller in the uncertain condition than the certain condition. This may have

accounted for why participants selected the optimal response less often in the uncertain condition (Herrnstein, 1961; Wasserman, Elek, Chatlosh, & Baker, 1993), without needing to appeal to uncertainty as a motivator. Indeed, Daw, O'Doherty, Dayan, Seymour, & Dolan (2006) have previously argued that, for designs where rewards are uncertain, simple choice rules that compare the relative value of different responses can provide an adequate account of exploration in choice behavior. As such, since average reward value differed markedly in the two conditions of Beesley et al.'s task, we cannot determine whether it was the uncertainty of the task, or the relative value of the responses, that motivated participants to make more suboptimal choices and show an increased level of attention to cues.

To address the issues highlighted above, we assessed exploratory behaviour in both choice and attention using a *contextual two-armed bandit* task (Schulz, Konstantinidis, & Speekenbrink, 2018), in which participants are given a free choice of two responses, and are told how many points they earned for making that response. The value of each response is determined by the *context* in which the decision is made, which is usually indicated through some explicit visual cue. Importantly, this task used a *limited*-feedback procedure: while participants were told the outcome of the response they made, they were not told what they would have received if they had made the alternative response. Hence, in this task we can monitor choice as a direct index of participants' exploration of the value of the different responses.

In the contextual two-armed bandit task, uncertainty can be manipulated by adjusting the variability of the payout produced by each response (Speekenbrink & Konstantinidis, 2015). Notably, this method allows us to introduce uncertainty without needing to change the mean value of each option. As an example, imagine a two-armed bandit task where selecting Response 1 always gives 10 points, and Response 2 gives between 4 and 16 points (drawing from a uniform distribution, represented by $U\{4, 16\}$). By introducing a distribution of

possible scores as the reward for response 2, there is now uncertainty in the specific outcome that will be produced when response 2 is chosen. However, responses 1 and 2 have the same objective mean value, so neither arm is (on average) better than the other.

The contextual two-armed bandit task therefore solves both of the issues we have raised with the Beesley et al. (2015) task. In their task, participants received full feedback on each trial as to what response was optimal on that trial, eliminating the need to actively explore different responses. As the proposed task has feedback regarding only the selected response, participants are required to intentionally explore the different responses to learn the best response to make for a given set of cues. Furthermore, by manipulating uncertainty as the *variability* associated with a reward (while holding average reward value constant), we can compare performance under certainty/uncertainty in which the two conditions are matched in terms of the overall objective reward that can be earned, and in the expected value of optimal and suboptimal response options on each trial.[2] If participants show greater suboptimal responding, and greater attention to cues, under conditions of uncertainty in this task, we can conclude that participants explore more due to uncertainty, and not the difference in reward value between optimal and suboptimal response options.

In summary, the current study aimed to assess whether uncertain learning situations increased exploration in attentional allocation in a limited-feedback task where participants were also required to explore the different response options as well as the usefulness of

[2] While we matched (objective) average reward in the certain/uncertain conditions, this did not necessarily mean that subjective value (utility) was exactly matched, since reward may be non-linearly related to utility. That said, between-conditions differences in utility were presumably smaller in this procedure than in previous research (Beesley et al., 2015), which had large between-conditions differences. Moreover, given the standard assumption of a negatively accelerated utility function (e.g., Kahneman & Tversky, 1979), the utility difference between optimal and suboptimal responses would be larger for our uncertain condition than the certain condition. To anticipate, our finding of smaller differences in choice behavior and attention for the uncertain condition therefore suggests the effect of uncertainty did not result from a between-conditions difference in utility.

different cues. In doing so, we can begin to unpack how choice behavior and attention may interact in the EE trade-off.

Method

Participants

Forty-four UNSW students (31 identified as female, 13 as male, age M = 19.4, SD = 2.2 years) participated for course credit. Participants were randomly allocated to the certain (n = 22) and uncertain (n = 22) conditions. Though a power analysis suggested 15 participants per group would be sufficient to observe a between-condition difference in proportion of attention (the smallest effect of interest, η_p^2 = .28, β = .99), we tested 44 participants to avoid the possibility that effect sizes were overestimated in Beesley et al. (2015). The experiment was approved by the UNSW Sydney Human Research Ethics Advisory Panel (Psychology). Three participants in the certain condition had no recorded fixations on cue stimuli on over 50% of trials, and an eye-tracker error meant one participant in the uncertain condition had no eye-tracking data. These participants were excluded from the eye-gaze analysis (we note that including these participants in eye-tracking analyses—or excluding them from analyses of choice behavior—did not alter the pattern of results or the conclusions drawn). MATLAB code for the experiment and all raw data are available via the Open Science Framework (OSF) at https://osf.io/kjz59/?view_only=6cfdb8ab656e44639ae28348f6749299.

We note for completeness that two further conditions were also run (n = 23 and 22); these conditions were a conceptual replication of the procedure of Beesley et al. (2015). The replication was successful; for brevity, a report on this replication can be found at the OSF link noted above, and on PsyArXiv at https://psyarxiv.com/uanmr.

Materials

The experiment was programmed using MATLAB with Psychophysics Toolbox extensions (Kleiner et al., 2007), presented on the 23-inch monitor (1920×1080 pixels, 60 Hz) of a Tobii TX-300 eye-tracker (sample rate 300 Hz). Participants used a chinrest positioned ~55 cm from the screen. Responses were "up" and "down" arrow keys. Four stylized tarot cards (504×360 pixels) were used as cue stimuli. The two cards shown on each trial were presented to the left and right of the screen (centers 1152 pixels apart). Feedback was presented centrally.

Design

Table 1 shows the relationship between cue-compounds and response outcomes for the two between-subject conditions. Uncertainty was manipulated via the variability associated with objective reward values for the different responses. For all participants, cues A and B were useful in predicting the best response to make on each trial. For participants in the certain condition, the presence of cue A meant response 1 yielded a high reward (15) on every trial, while response 2 yielded a low reward (10) on every trial. The opposite was true for cue B. In the uncertain condition, cues A and B were useful in predicting which response would yield the higher reward on the majority of trials (the optimal response). However, while the mean values of the high-value and low-value rewards across trials were 15 and 10 points respectively, the specific reward on each trial was drawn from a uniform distribution with a range of 13 points. For example, if the participant made the optimal response, the reward they received on that trial ranged between 9 and 21 points ($U\{9, 21\}$). If they made the alternate response, the reward ranged from 4 to 16 points ($U\{4, 16\}$). In both conditions, cues X and Y were non-predictive: the presence of either of these cues provided no information on the outcome that would be obtained by either response.

Procedure

Participants were instructed they would play a card game in which they viewed two cards and betted “up” or “down”. They could use feedback to learn how the cards related to responses and rewards, to maximize their accumulated points. Participants were informed that the top two performers (based on accumulated points) would receive $20.

On each trial, a central black fixation cross appeared for 1 second, after which the two cards appeared. Participants had unlimited time to view the stimuli and make a response. Feedback was then presented for 1.5 seconds, showing the number of points earned and the total points accumulated. The next trial began after an inter-trial interval of 1.5 seconds. Participants were given three self-timed breaks, spaced evenly throughout the task.

The experiment consisted of 256 trials, split into 8 epochs of 32 trials. Every 8 trials, each compound of cues (AX, AY, BX, BY) was presented twice, with the left-right spatial positioning of the two cues counterbalanced across these two presentations. Trial order within these 8 trials was random, with the constraint that the same compound could not be presented on consecutive trials. The sequence of rewards for each response was generated at the start of the procedure. We constrained the sequence to ensure that both groups had the same average reward for each response (i.e., 15 for the optimal response, 10 for the suboptimal response). This was achieved by assessing the mean value of the generated sequence in a moving window for every consecutive set of 8 trials (i.e., trials 1-8, 2-9, etc). If the mean value of those eight trials deviated by more than three points from the underlying mean, the entire sequence of trials was resampled and assessed again from the beginning.

Results

Choice behavior

Figure 1 shows the proportion of optimal responses made in each 32-trial epoch for each condition. Participants learned to make more optimal responses over time, with a clear difference in optimal response rate across the two groups. An ANOVA, with a within-subjects factor of epoch and between-subjects factor of uncertainty, confirmed this interpretation. There was a main effect of epoch, $F(7,294) = 22.46$, $p < .001$, $\eta_p^2 = .348$, a main effect of uncertainty, $F(1,42) = 9.68$, $p < .001$, $\eta_p^2 = .187$, and a significant interaction between epoch and uncertainty, $F(7,294) = 2.61$, $p = .013$, $\eta_p^2 = .058$, with optimal responding increasing at a faster rate for participants in the certain condition compared to participants in the uncertain condition.

Eye gaze

Figure 2 shows the proportion of trial-time spent looking at the predictive and non-predictive cues in each epoch for each condition. We analyzed these data using ANOVA with within-subjects factors of epoch and predictiveness (predictive vs. non-predictive cue), and a between-subjects factor of uncertainty (certain vs. uncertain). We found that fixation time decreased across epochs, $F(7,266) = 26.69$, $p < .001$, $\eta_p^2 = .413$, was significantly higher for participants in the uncertain condition than the certain condition, $F(1,38) = 12.01$, $p = .001$, $\eta_p^2 = .240$, and this difference increased across epochs, $F(7, 266) = 4.84$, $p < .001$, $\eta_p^2 = .113$. Greater fixation time was devoted to predictive over non-predictive cues, $F(1,38) = 42.83$, $p < .001$, $\eta_p^2 = .530$. This difference increased across epochs, $F(7,266) = 5.12$, $p < .001$, $\eta_p^2 = .119$, and was greater in the certain condition than the uncertain condition, $F(1, 38) = 7.73$, $p = .008$, $\eta_p^2 = .169$. Despite this, the effect of predictiveness was significant in both the certain, $F(1, 18) = 29.92$, $p < .001$, $\eta_p^2 = .624$, and the uncertain conditions,

$F(1, 20) = 11.39$, $p = .003$, $\eta_p^2 = .363$. The three-way interaction in the omnibus ANOVA was not significant, $F(7, 266) = 1.98$, $p = .058$, $\eta_p^2 = .050$.

Discussion

Our data suggest that uncertainty promotes an exploratory profile of behavior in both choice and attention. Taking first the choice data, we found that participants tended to issue the optimal response at a reduced rate under conditions of uncertainty compared to conditions of certainty. The limited-feedback procedure used here presents a situation in which changing from the optimal to the suboptimal response can be taken to reflect an exploratory response, where the participant seeks to gain new knowledge about response payouts. Furthermore, unlike in Beesley et al. (2015), where average reward payouts differed substantially between certain and uncertain conditions, the current study's objective mean reward payouts were matched across conditions. Our findings therefore suggest that, contrary to previous claims (Daw et al., 2006) uncertainty can drive exploration (biasing the EE trade-off) even in the absence of marked differences in the expected rewards for different responses.

Turning to attention, participants in the uncertain condition spent longer looking at cues overall (as a proportion of trial time) compared to participants in the certain condition (see also Beesley et al. 2015; Easdale et al., 2017; Luque, Vadillo, Le Pelley, & Beesley, 2017). Participants also spent a greater proportion of trial-time fixating on predictive cues over non-predictive cues: this greater attention to more informative cues suggests 'attentional exploitation'. This attentional exploitation was particularly pronounced in the certain condition, with strong preferences towards the predictive cue over the non-predictive cue. In contrast, participants in the uncertain condition maintained a high level of attention to both the predictive and non-predictive cues, suggesting an exploratory mode of attention.

Our findings fit into an emerging literature on the interaction of learning and attention in reinforcement learning. It has been suggested that attention might be used in reinforcement learning as a mechanism to solve the problem of stimulus dimensionality (Niv et al., 2015). This describes how, when a learning agent is in a complex environment, it must "prune out" uninformative stimulus dimensions from attentional processing, while still attending to relevant stimulus dimensions. While participants in our study did appear to prune out non-predictive cues under conditions of certainty, the opposite pattern was observed in uncertainty, with all cues receiving *increased* attention. We propose that this indiscriminate increase in attention to cues reflects *attentional exploration*, while the pruning out of non-predictive cues from attention reflects *attentional exploitation*.

Our findings extend previous work in the learning literature. For example, research in category learning has used eye-tracking to show that participants can learn to attend to relevant stimulus dimensions and ignore irrelevant dimensions (e.g., Rehder & Hoffman, 2005; see also Le Pelley et al., 2011), akin to the attentional exploitation demonstrated here (though these previous studies did not investigate the effect of manipulating environmental uncertainty). More recently, Braunlich and Love (2019) have created the Sampling Emergent Attention model of category learning, which postulates that participants can choose to sample stimulus dimensions in either an exploitative manner (sampling known relevant dimensions), or in an exploratory manner (sampling dimensions that maximize information gain). Notably, studies of category learning typically use *full-feedback* procedures: regardless of the response made, participants are told what would have been the 'correct' response on each trial. Our data suggest that similar processes may operate under the limited-feedback conditions more typical of reinforcement learning. This procedure also allowed us to show, for the first time, that uncertainty influences the EE trade-off in both attention and choice behavior within the same task.

To conclude, the current data reflect an intuitive relationship between attention and learning. If an agent is satisfied by the reward it will receive from a response, and does not expect to gain new information through exploring other responses, it should not expend effort attending to stimuli that are not useful to its immediate decision. In contrast, if an agent is less sure about the values of its available responses, it should widen its attention in order to gather information that may be helpful in making better decisions in the future. It is clear that attention is an important component of the behavior within situations that reflect an EE trade-off, and reinforcement learning more broadly.

Open Practices Statement

The data and materials for the current study are available at https://osf.io/kjz59/?view_only=6cfdb8ab656e44639ae28348f6749299. The current study was not preregistered.

References

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

Braunlich, K., & Love, B. C. (in press). Occipitotemporal representations reflect individual differences in conceptual knowledge. *Journal of Experimental Psychology: General.*

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.

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

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

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4*(3), 267–272.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica, 47,* 263-291.

Kleiner, M., Brainard, D., Pelli, D., (2007). What's new in Psychtoolbox-3. *Perception, 36,* 1.

Knox, W. B., Otto, A. R., Stone, P., & Love, B. C. (2012). The nature of belief-directed exploratory choice in human decision-making. *Frontiers in Psychology*, *2*, 1–12.

Le Pelley, M. E., Beesley, T., & Griffiths, O. (2011). Overt Attention and Predictiveness in Human Contingency Learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*, 220–229.

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

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

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.

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*, 8145–8157.

Schulz, E., Konstantinidis, E., & Speekenbrink, M. (2018). Putting bandits into context: How function learning supports decision making. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 44, 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*, 233–250.

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

Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating Causal Relations: Role of Probability in Judgments of Response-Outcome Contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 174–188.

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.

Tables and Figures

Table 1: *Rewards for making response 1 (R1) and response 2 (R2) for the indicated compound in the certain and uncertain conditions.*

Condition	Cue compound	Reward (R1)	Reward (R2)
Certain	AX	15	10
	AY	15	10
	BX	10	15
	BY	10	15
Uncertain	AX	$U\{9, 21\}$	$U\{4, 16\}$
	AY	$U\{9, 21\}$	$U\{4, 16\}$
	BX	$U\{4, 16\}$	$U\{9, 21\}$
	BY	$U\{4, 16\}$	$U\{9, 21\}$

Figure 1: Mean proportion of optimal responses for participants in the certain and uncertain conditions. Error bars represent standard error of the mean.

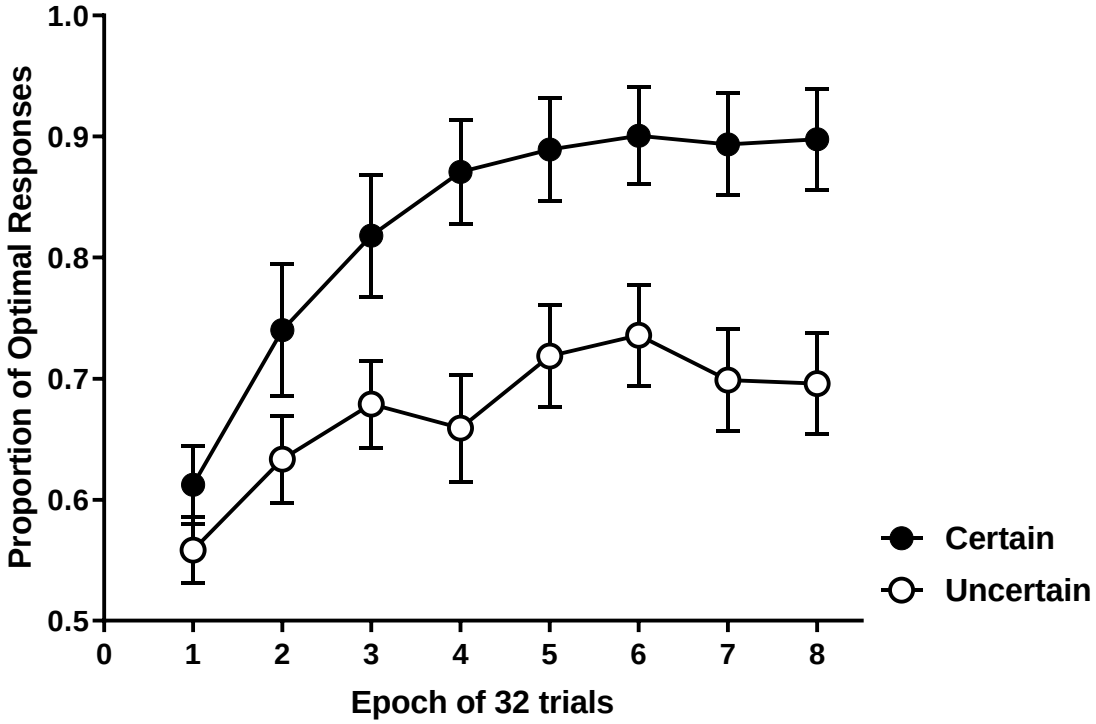

Figure 2: Mean proportion of trial time spent fixating on predictive (P) and non-predictive (NP) cues, for participants in the certain and uncertain conditions. Error bars represent standard error of the mean.