Under pressure: The influence of time limits on human exploration

Charley M. Wu[1,*]**(cwu@mpib-berlin.mpg.de), Eric Schulz**[2,*]**,**
Kimberly Gerbaulet[1,3,*]**, Timothy J. Pleskac**[1,4]**, & Maarten Speekenbrink**[5]
[1]Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany
[2]Department of Psychology, Harvard University, Cambridge, Massachusetts, USA
[3]Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany
[4]Department of Psychology, University of Kansas, Lawrence, Kansas
[5]Department of Experiment Psychology, University College London, London, UK
[*]Contributed equally to this work.

Abstract

How does time pressure influence attitudes towards uncertainty? When time is limited, do people engage in different exploration strategies? We study human exploration in a range of four-armed bandit tasks with different reward distributions and manipulate the available time for each decision (limited vs. unlimited). Through multiple behavioral and model-based analyses, we show that reactions towards uncertainty are influenced by time pressure. Specifically, participants seek out uncertain options when time is unlimited, but avoid uncertainty under time pressure. Moreover, larger relative differences in uncertainty between options slowed down reaction times and dampened the drift rate of a linear ballistic accumulator model. These results shed new light on the differential effect of uncertainty and time pressure on human exploration.

Keywords: Exploration-exploitation; Uncertainty; Time Pressure; Directed Exploration; Multi-armed Bandits

Introduction

Searching for rewards requires navigating the exploration-exploitation dilemma: Should one exploit options known to produce high rewards, or explore lesser known options to gain information that could potentially lead to even higher rewards? Because optimal solutions (Gittins, 1979) are generally intractable in realistic settings, practical solutions usually rely on heuristics (Auer, Cesa-Bianchi, & Fischer, 2002), which can be classified as directed exploration, random exploration, or both.

Directed exploration is often implemented using an exploration bonus that inflates the expected value of an option proportional to the estimated uncertainty, to encourage the exploration of uncertain options. Whereas earlier studies produced mixed evidence for the use of exploration bonuses in human reinforcement learning (Daw, O'doherty, Dayan, Seymour, & Dolan, 2006), there is now an increasing amount of evidence for directed exploration in vast problem spaces (Wu, Schulz, Speekenbrink, Nelson, & Meder, 2018), planning (Wilson, Geana, White, Ludvig, & Cohen, 2014), dynamic decision making (Knox, Otto, Stone, & Love, 2012), and simple two-armed bandit tasks (Gershman, 2018).

Unlike directed exploration, *random exploration* increases choice stochasticity in accordance to the agent's uncertainty about the value of available actions (Speekenbrink & Konstantinidis, 2015). One recent theory proposed that random and directed exploration can be dissociated, where the balance is influenced by the total and relative uncertainty of available options (Gershman, in press). If there are multiple options with similar expected rewards, directed exploration makes an option more likely to be sampled when its uncertainty is higher relative to the other options (Schulz & Gershman, 2019). We make use of this effect by studying how patterns of decision making and exploration are affected by both uncertainty and expected reward in a four-armed bandit task. Compared to previously studied two-armed bandit tasks, the richer set of options makes exploration more pertinent and observable over more trials. Crucially, we manipulate the presence or absence of time pressure to gain insights into the cognitive processes underlying exploration. If directed exploration is a reasoned and controlled process, which requires taking the uncertainties of each options into account, then time pressure may limit the capacity for directed exploration.

As predicted, we find that participants are more likely to sample options with high relative uncertainty in the absence of time pressure. However, when we impose time pressure by limiting the allowed decision time to under 400 milliseconds, we find that relative uncertainty reduces the probability that an option is chosen. Additionally, relative uncertainty slows down reaction times more strongly and dampens the evidence accumulation process more heavily under time pressure. In other words, time pressure moderates the effect of environmental uncertainty, such that risk-seeking behavior arising through directed exploration transforms into risk-aversion under time pressure. These results enrich our understanding of human exploration strategies under changing task demands.

Experiment

Participants and Design. We recruited 99 participants (36 female, aged between 21 and 69 years; M=34.82; SD=10.1) on Amazon Mechanical Turk (requiring 95% approval rate and 100 previously approved HITs). Participants were paid \$3.00 for taking part in the experiment and a performance contingent bonus of up to \$4.00 (calculated based on the performance of one randomly selected round). Participants spent 13.0 ± 5.6 minutes on the task and earned $\$5.87 \pm \$0.91$ in total. The study was approved by the Ethics Committee of the Max Planck Institute for Human Development.

We used a 2×4 within-subject design to examine how the presence or absence of time pressure and the payoff structure of the task (see Fig. 1b and Tab. 1) influenced choices and reaction times. In total, the experiment consisted of 40 rounds

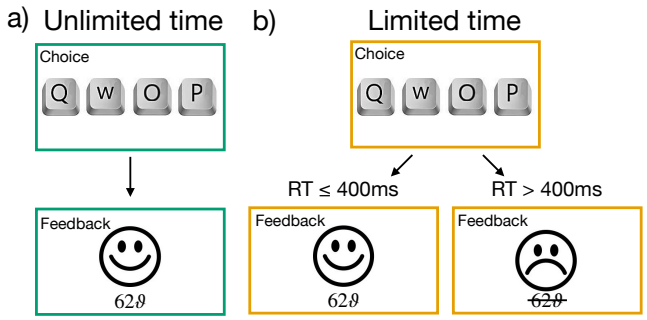

Figure 1: **Experimental design.** We used a four-armed bandit task where each option was randomly mapped to the Q, W, O, and P keys on the keyboard. **a**) In unlimited time rounds, participants could take as long as they wanted to make each selection and received positive feedback (happy face) and were shown the value of the acquired payoff. **b**) In limited time rounds, participants were only given 400 ms to make each selection. If they exceeded the time limit, they would forgo earning any rewards, and received negative feedback (sad face) along with the value of the payoff they could have earned (crossed out).

with 20 trials each. In each round, a condition was sampled (without replacement) from a pre-randomized list, such that each combination of time pressure and payoff structure was repeated five times, with a total of 100 trials in each.

Materials and Procedure. Participants were required to complete three comprehension questions and two practice rounds (one with unlimited time and one with limited time) consisting of 5 trials each before starting the experiment. Each of the 40 rounds was presented as a four-armed bandit task, where the four options were randomly mapped to the $[Q,W,O,P]$ keys on the keyboard (Fig. 1). Selecting an option by pressing the corresponding key yielded a reward sampled from a normal distribution, where the mean and variance was defined by the round's payoff structure (Fig. 2a and Tab. 1). Participants completed 20 trials in each round and were told to acquire as many points as possible.

Before starting a round, participants were informed whether it was an unlimited or a limited time round. In unlimited time rounds, participants could spend as much time as they needed to reach a decision, upon which they were given feedback about the obtained reward (displayed for 400 ms) before continuing to the next trial (Fig. 1a). In limited time rounds, participants were instructed to decide as fast as possible. If a decision took longer than 400 ms, they forfeited the reward they would have earned (presented to them as a crossed-out number with an additional sad smiley; Fig. 1b). We used the same inter-trial period of 400 ms to display feedback about obtained rewards in both limited and unlimited time rounds.

We applied a random shifting of rewards across rounds (i.e., different maximum reward) to prevent participants from immediately recognizing when they had chosen the optimal option. For each round, we sampled a value from a uniform distribution $\mathcal{U}(30,60)$, which was then added to the rewards. Together with random shifting, we also truncated rewards such that they were always larger than zero. In order to convey intuitions about the random shift of rewards, payoffs were presented using a different fictional currency in each round (e.g., ß, Þ, ϑ), such that the absolute value was unknown, but higher were always better.

Table 1: Payoff Conditions

Payoff Conds	Means (μ)	Variances (σ^2)
IGT	$[-10,-10,10,10]$	$[10,100,10,100]$
Low Var	$[-10,-\frac{1}{3},\frac{1}{3},10]$	$[10,10,10,10]$
High Var	$[-10,-\frac{1}{3},\frac{1}{3},10]$	$[100,100,100,100]$
Equal Means	$[0,0,0,0]$	$[10,40,70,100]$

At the end of each round, participants were given feedback about their performance in terms of the bonus they would gain (in USD) if this was the round selected for determining the bonus. The bonus was calculated as a percentage of the total possible performance, raised to the power of 4 to accentuate differences in the upper range of performance:

$$\text{Bonus} = \left(\frac{\text{total reward gained}}{\text{mean reward of best option} \times 20 \text{ trials}}\right)^4 \times \$4.00$$

Payoff conditions We used four different payoff conditions as a within-participant manipulation (Tab. 1 and Fig. 2a). Each payoff condition specified the mean μ_i and variance σ_i^2 of the reward distribution $R_i \sim \mathcal{N}(\mu_i,\sigma_i^2)$ for each option i. Each distribution was randomly mapped to one of the four $[Q,W,O,P]$ keys of the keyboard in each round. The Iowa Gambling Task (IGT) is a classic design that has been related to a variety of clinical and neurological factors affecting decision-making (Yechiam, Busemeyer, Stout, & Bechara, 2005; Bechara, Damasio, Damasio, & Anderson, 1994). We implemented a reward condition inspired by the IGT such that there are two high and two low reward options, with a low and high variance version of each. We also constructed two conditions with equally spaced means, but with either uniformly low variance or uniformly high variance. Lastly, the equal means condition had identical means and gradually increasing variance, such that we can observe the influence of uncertainty independent of mean reward.

Behavioral Results

Participants acquired higher rewards in the unlimited than in the limited time condition (Fig. 2b; $t(98) = 3.1$, $p = .002$, $d = 0.3$, $BF = 10$). Participants also improved over trials, signified by an average correlation between trial and rewards (Spearman's $\rho(98) = 0.16$, $p < .001$, $BF > 100$). This correlation did not differ between limited and unlimited time rounds ($t(98) = -1.3$, $p = .196$, $d = 0.1$, $BF = .25$).

We also compared performance across payoff conditions. This is possible, since all games had the same expected reward under the assumption of a random sampling strategy. We found that participants performed better in the IGT-like condition than in the low variance condition ($t(98) = 3.2$, $p = .002$, $d = 0.3$, $BF = 14$). We see an even larger difference when comparing the low variance and high variance conditions, which had the same means but different levels of

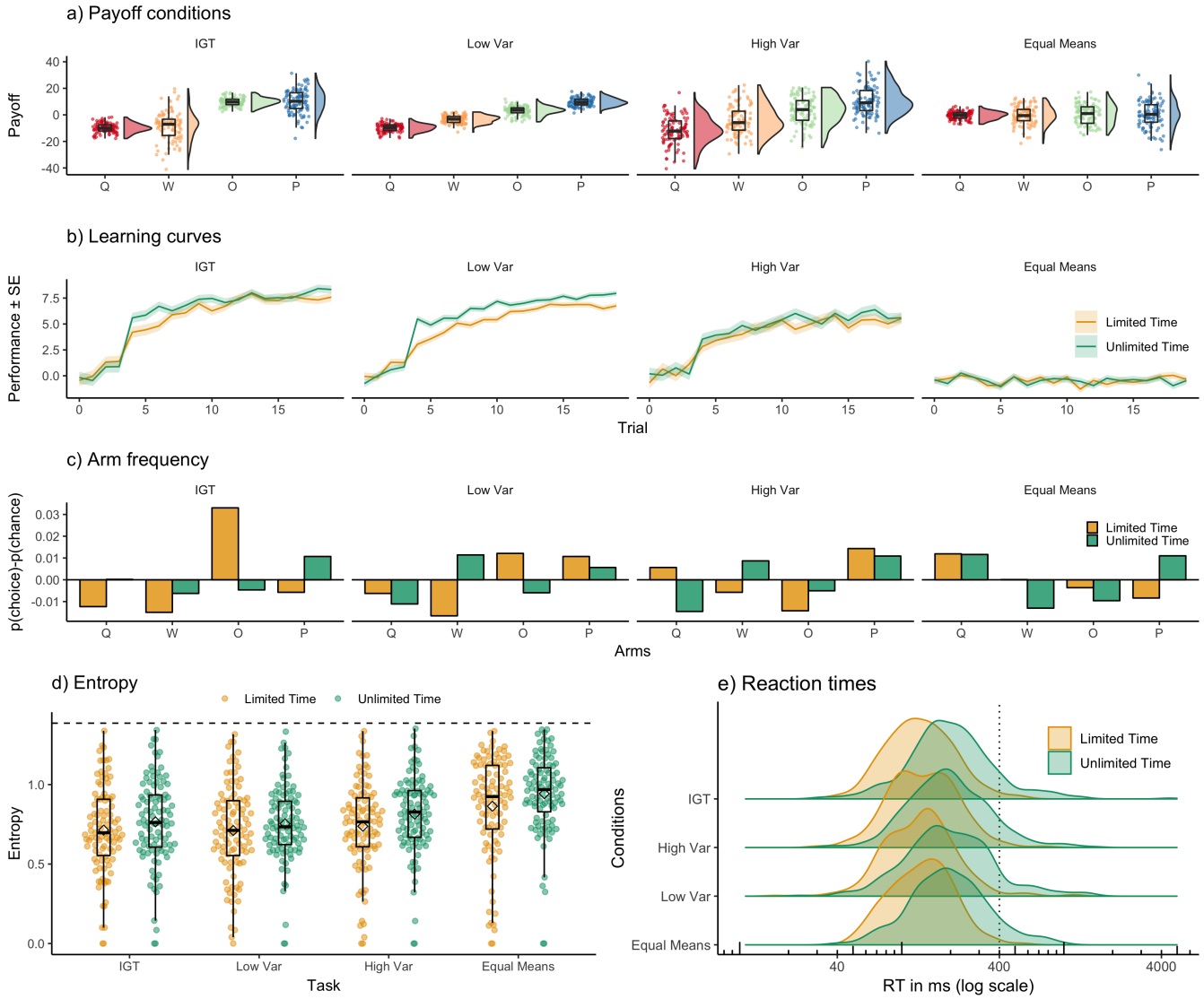

Figure 2: **Payoff conditions and behavioral results**. **a**) Four different payoff conditions were combined with either limited or unlimited time rounds to create 8 different scenarios. Each condition specifies a normal payoff distribution for each option; the means and variances are shown in Table 1. Each dot represents a randomly drawn payoff, while the Tukey boxplots and half violin plots show the distribution for 100 simulated draws. Note that rewards were randomly shifted in each round by adding a constant $\sim \mathcal{U}(30,60)$ to all payoffs. **b**) Learning curves of average participant performance (using unshifted rewards) over trials by payoff condition. Ribbons indicate standard error. **c**) Choice proportions (normalized for chance) for each option, mapped to the canonical ordering shown in panel a). **d**) The entropy of choices in each round, where higher entropy corresponds to more diverse choices and the dotted line indicates random chance (i.e., playing each arm with equal probability). Each dot represents a participant, and overlaid are Tukey boxplots with the diamond indicating the group mean. **e**) Distributions of reaction times in milliseconds (ms) and shown on a log scale. The vertical dotted line indicates the time limit (400 ms) of the limited time condition

risk and uncertainty. Participants performed substantially better in the low variance condition than the high variance condition ($t(98) = 6.2$, $p < .001$, $d = 0.6$, $BF > 100$). Thus, higher variance increased the difficulty of the task. Lastly, participants performed better in the high variance than in the equal means task ($t(98) = 25.5$, $p < .001$, $d = 2.6$, $BF > 100$), which is intuitive since improvement is not possible if all arms have the same mean reward.

Choice proportions. Figure 2c shows the proportion of choices, which illustrates differences across time conditions. We used a Bayesian mixed-effects logistic regression and found that in the IGT condition, participants chose the high reward-low variance option (indicated as 'O' in Fig. 2c) less frequently in the unlimited time than in the limited time condition ($\hat{\beta} = -.22$, 95% Highest Posterior Density (HPD) interval: $[-.28, -.15]$, $BF > 100$)[1].

Additionally, we also find differences across time-pressure conditions in the Equal Means task, where participants selected the highest variance option ('P') more frequently in the unlimited time condition $\hat{\beta} = .11$, 95% HPD: $[.05, .17]$,

[1]We use Bridge sampling (Gronau, Singmann, & Wagenmakers, 2017) to approximate the Bayes Factor by comparing against an intercept-only null model (i.e., without time pressure as a predictor).

$BF = 15$). This illustrates a shift in preferences away from uncertain options when time pressure is introduced. Whereas participants tend to be risk-seeking and choose highly uncertain options under unlimited time, they become more risk-averse and choose them less often under time pressure.

We also calculated the Shannon entropy of participants' choices in each round (Fig. 2d), where higher entropy corresponds to higher diversity of choices and the maximal entropy strategy would be to choose each option an equal number of times (indicated by the dotted line). Averaged across participants, we find higher choice entropy (i.e. more diversity in choice) under unlimited time than limited time ($t(98) = 4.1$, $p < .001$, $d = 0.4$, $BF > 100$). This further strengthens the evidence for reduced exploration under time pressure, since we find a lower diversity of choices.

Reaction times. Figure 2d shows reaction times. Unsurprisingly, participants responded faster in the limited time than in the unlimited time conditions (comparing RTs in logms: $t(98) = 9.7$, $p < .001$, $d = 1.0$, $BF > 100$). There were no differences across payoff conditions ($F(3,95) = 0.12$, $p = .951$, $BF = 0.01$).

Model-Based Analyses

In order to model learning and decision making in our task, we use a *Bayesian mean tracker* (BMT) as a reinforcement learning model for estimating rewards and uncertainties, which are then updated based on prediction error. The BMT is a variant of a Kalman filter, but assumes a time-invariant reward distribution (as is the case in our experiment) instead of a dynamically changing one. Both models use an updating rule based on prediction error, and have been described as a Bayesian extension of the classic Rescorla-Wagner model of associative learning (Gershman, 2015). Variants of the BMT have been used to describe human behavior in a variety of multi-armed bandit and decision-making tasks (Gershman, 2018, in press; Yu & Dayan, 2003; Schulz, Konstantinidis, & Speekenbrink, 2015; Dayan, Kakade, & Montague, 2000; Speekenbrink & Konstantinidis, 2015).

The BMT learns a posterior distribution over the mean reward μ_j for each option j. Rewards are assumed to be normally distributed with a known variance but unknown mean. The prior distribution of the mean is also a normal distribution. This implies that the posterior distribution for each mean is also a normal distribution:

$$p(\mu_{j,t}|\mathcal{D}_{t-1}) = \mathcal{N}(m_{j,t}, v_{j,t}) \tag{1}$$

where $\mathcal{D}_{t-1}$ denotes the previously observed rewards for all options. For a given option j, the posterior mean $m_{j,t}$ and variance $v_{j,t}$ are only updated when it has been selected at trial t:

$$m_{j,t} = m_{j,t-1} + \delta_{j,t} G_{j,t} [y_t - m_{j,t-1}] \tag{2}$$

$$v_{j,t} = [1 - \delta_{j,t} G_{j,t}] v_{j,t-1} \tag{3}$$

where $\delta_{j,t} = 1$ if option j is chosen on trial t, and 0 otherwise. Additionally, y_t is the observed reward at trial t, and $G_{j,t}$ is defined as:

$$G_{j,t} = \frac{v_{j,t-1}}{v_{j,t-1} + \theta_\varepsilon^2} \tag{4}$$

where θ_ε^2, referred to as the error variance, is the variance of the rewards around the mean. For our model-based analysis, we set the error variance to 1 (which led to competitive task performance in prior simulations).

Intuitively, the estimated mean of the chosen option $m_{j,t}$ is updated based on prediction error, which is the difference between the observed reward y_t and the prior expectation $m_{j,t-1}$, multiplied by learning rate $G_{j,t} \in [0,1]$. At the same time, the estimated variance $v_{j,t}$ of the chosen option is reduced by a factor $1 - G_{j,t}$. The error variance (θ_ε^2) can be interpreted as an inverse sensitivity, where smaller values result in more substantial updates to the mean $m_{j,t}$, and larger reductions of uncertainty $v_{j,t}$. We set the prior mean to $m_{j,0} = 45$ and the prior variance to $v_{j,0} = 55$ based on the expectation across payoff conditions.[2]

Results

We followed Gershman (in press) and generated predictions from the BMT by feeding in a participant's observations on a particular round until time t, and then predicting the mean and standard deviation for each option at time point $t+1$. We used the resulting predictions of rewards and uncertainties to conduct three model-based analyses of choices, reaction times, and evidence accumulation.

Choices. In our first analysis, we assessed how the predicted mean and uncertainty of an option affected the likelihood of it being chosen on each trial (estimated separately for limited and unlimited time conditions). We applied hierarchical Bayesian inference to estimate the parameters of a softmax policy, under the assumption that a participant's choice on each trial is influenced by both the predicted mean and uncertainty of an option, where each participant's parameters are assumed to be jointly normally distributed. The probability of choosing option j on trial t is a softmax function of its decision value $Q_{j,t}$:

$$P(C_t = j) = \frac{\exp(Q_{j,t})}{\sum_{k=1}^{4} \exp(Q_{k,t})}. \tag{5}$$

The decision value $Q_{j,t}$ is a linear function of the estimated mean $m_{j,t}$ and uncertainty $\sqrt{v_{j,t}}$ (estimated as a standard deviation) of each option according to the BMT:

$$Q_{j,t} = \beta_1 m_{j,t} + \beta_2 \sqrt{v_{j,t}}. \tag{6}$$

Formally, we assume that the β-coefficients for each participant $\beta_i = (\beta_{1,i}, \beta_{2,i})$ are drawn from a normal distribution

$$\beta_i \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2), \tag{7}$$

[2] We use the shifted reward values that were observed by participants, where the means in each condition were centered on 0 and shifted by $\mathcal{U}(30,60)$.

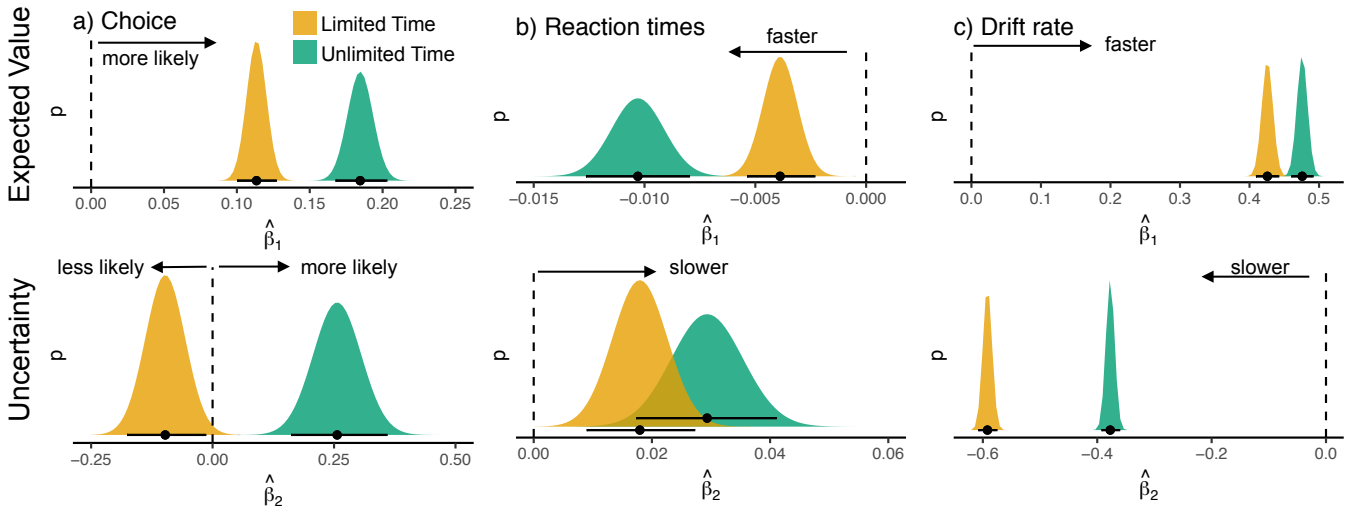

Figure 3: **Posterior parameter estimates. a**) Effects of BMT predicted mean rewards ($\hat{\beta}_1$) and uncertainties ($\hat{\beta}_2$) on an option's probability of being chosen, estimated by a hierarchical Bayesian softmax regression. **b**) Influence of BMT means ($\hat{\beta}_1$) and uncertainties ($\hat{\beta}_2$) on participant response times, estimated by a hierarchical Bayesian linear regression. **c**) Influence of BMT means ($\hat{\beta}_1$) and uncertainties ($\hat{\beta}_2$) on drift rates in a Bayesian Linear Ballistic Accumulator model. In all plots, the vertical dashed line indicates an effect of 0, while the black dot indicates the mean effect and confidence intervals show the 95% highest posterior density (HPD).

and we estimate the group-level mean μ_β and variance over participants σ^2_β. We used the following priors on the group-level parameters:

$$\mu_\beta \sim \mathcal{N}(0, 100) \tag{8}$$

$$\sigma_\beta \sim \text{Half-Cauchy}(0, 100) \tag{9}$$

In each time condition, we arrive at group-level parameter estimates describing how expected rewards (β_1) and uncertainty (β_2) influence choice probability under the softmax policy.

We estimated the hierarchical model using Hamiltonian Markov chain Monte Carlo sampling with `PyMC3` (Salvatier, Wiecki, & Fonnesbeck, 2016). The results (Fig 3a) show that the expected value of an option increased choice probability for both the limited time ($\hat{\beta}_1 = .11$, 95% HPD: $[.10, .13]$) and the unlimited time conditions ($\hat{\beta}_1 = .19$, 95% HPD: $[.17, .2]$). Options estimated to have higher expected rewards were more likely to be chosen in both conditions, with a stronger effect in the unlimited time conditions.

Notably, we found contrasting effects of uncertainty on choice probability. In the unlimited time conditions, uncertainty had a positive effect on choice probability ($\hat{\beta}_2 = .26$, 95% HPD: $[.16, .36]$). This replicates previous findings reported in two-armed bandit tasks without time pressure (Gershman, 2018, in press). However, uncertainty had a negative effect on choice probability in the limited time condition ($\hat{\beta}_2 = -.10$, 95% HPD: $[-.18, -.02]$). Thus, whereas participants sought out uncertain options in the unlimited time condition, they shunned uncertain options in the limited time condition.

Reaction Time. Our second analysis looked at how the estimated means and uncertainties of options influenced reaction times. We normalized the BMT predictions of mean reward and uncertainty by calculating the difference between the chosen option and the average of the unchosen options on each trial. Thus, positive values indicate that expected reward/uncertainty are relatively larger than those of the unchosen options. We regressed these normalized means and uncertainties onto participant log reaction times[3] in a hierarchical Bayesian linear regression, using the same priors over the β-coefficients as before (Eq. 9).

The resulting posterior parameter estimates (Fig. 3b) show that participants were faster at choosing options with relatively higher expected reward in both conditions, but with a stronger effect in the unlimited ($\hat{\beta} = -.01$, 95% HPD: $[-.013, -.008]$) than in the the limited time condition ($\hat{\beta} = -.004$, 95% HPD: $[-.005, -.002]$). Furthermore, participants were slower at choosing options with higher relative uncertainty in both the limited ($\hat{\beta} = .02$, 95% HPD: $[.01, .03]$) and the unlimited conditions ($\hat{\beta} = .03$, 95% HPD: $[.02, .04]$). Thus, whereas higher relative value made participants act faster, higher relative uncertainty slowed them down. This differs from previous findings using two-armed bandits (Gershman, in press), which showed higher relative uncertainty makes participants choose faster.

Evidence Accumulation. In our third analysis, we used the Linear Ballistic Accumulator (LBA; Brown & Heathcote, 2008) to model choices and reaction times simultaneously. This model assumes that choices are the result of a process in which evidence for each option is accumulated continuously over time, and that option is chosen for which the accumulated evidence first exceeds a set decision threshold.

Formally, the LBA assumes that, after an initial period of non-decision time τ, evidence for an option j on trial t accumulates at a rate of $v_{j,t}$, starting from an initial evidence

[3] 1 ms was added to each RT to avoid $\log(0)$. Additionally, RTs were truncated at 5000 ms.

level $p_{j,t} \sim \mathcal{U}(0, A)$. Evidence accumulates for each option j until a threshold b is reached. We follow the Bayesian implementation proposed by Annis, Miller, and Palmeri (2017) and assume that the priors for the drift rates stem from truncated normal distributions

$$v_{j,t} \sim \mathcal{N}(2,1) \in (0,\infty). \tag{10}$$

Additionally, we assume a uniform prior on non-decision time

$$\tau \sim \text{Uniform}(0,1), \tag{11}$$

and a truncated normal prior on the maximum starting evidence

$$A \sim \mathcal{N}(0.5,1) \in (0,\infty). \tag{12}$$

Finally, we reparameterized the model by shifting b by k units away from A, and put a truncated normal distribution as the prior on the resulting relative threshold k:

$$k \sim \mathcal{N}(0.5,1) \in (0,\infty). \tag{13}$$

We estimated the LBA parameters for each participant in every round using No-U-Turn Hamiltonian MCMC (Hoffman & Gelman, 2014), with reaction times truncated at 5000 ms. Participants had higher mean drift rates under limited time compared to unlimited time ($t(98) = 7.1$, $p < .001$, $d = 0.7$, $BF > 100$), consistent with the need to arrive at decisions more quickly. Participants in the limited time conditions also had shorter non-decision times τ ($t(98) = -4.6$, $p < .001$, $d = 0.5$, $BF > 100$), less maximum starting evidence A ($t(98) = -7.8$, $p < .001$, $d = 0.8$, $BF > 100$), and lower relative thresholds k ($t(98) = -5.2$, $p < .001$, $d = 0.5$, $BF > 100$), compared to participants in the unlimited time conditions. Thus, our LBA results confirm the intuition that participants thought more carefully about different options given unlimited time.

We then regressed the BMT predictions of relative expected reward and relative uncertainty for each option onto its estimated drift rate using a Bayesian linear regression. The result of this analysis revealed that the relative expected value of an option had a positive effect on drift rate for both the limited ($\hat{\beta} = .43$, 95% HPD: $[.41, .44]$; see Fig. 3c) and unlimited time conditions ($\hat{\beta} = .48$, 95% HPD: $[.46, .49]$), with a stronger effect in the latter. Conversely, relative uncertainty had a negative effect on drift rate, which was larger in magnitude for the limited ($\hat{\beta} = -.59$, 95% HPD: $[-.61, -.58]$) than for the unlimited time conditions ($\hat{\beta} = -.38$, 95% HPD: $[-.39, -.36]$). Thus, the behavioral patterns in Figure 2b suggest that uncertainty reduced the rate of evidence accumulation, with a stronger effect under time pressure than in the unlimited time conditions.

Discussion and Conclusion

How do people explore uncertain options under time pressure? We investigated this question using several variants of a four-armed bandit task with continuous rewards, while manipulating the available decision time to be either unlimited or limited to less than 400 ms.

Our models showed that higher relative uncertainty made an option more likely to be chosen in the absence of time pressure. This matches previous findings showing evidence for an exploration bonus consistent with directed exploration (Gershman, in press). However, putting participants under time pressure inverted this relationship, and caused uncertainty to reduce the probability that an option was chosen. Thus, the uncertainty bonus found in standard multi-armed bandit tasks can turn into an uncertainty penalty when people are under time pressure. This is similar to findings from description-based gambles, where time pressure increased risk aversion (Nursimulu & Bossaerts, 2013).

We also found that relative uncertainty slowed down choices and dampened evidence accumulation. These results suggest that uncertainty can have reversible effects on preference: sometimes people seek out uncertainty, and sometimes they actively avoid it. Both of these cases suggest people track uncertainty in their expectations, and that uncertainty feeds into the decision-making process. This is similar to what has been observed in tasks that directly elicit confidence judgments (Boldt, Blundell, & De Martino, 2017; Stojic, Schulz, Analytis, & Speekenbrink, 2018; Schulz, Wu, Ruggeri, & Meder, 2018; Wu, Schulz, Garvert, Meder, & Schuck, 2018), while previous work has shown that changing the context from only gains to adding risky options can also cause a shift from actively seeking uncertainty to avoiding it (Schulz, Wu, Huys, Krause, & Speekenbrink, 2018).

Our results provide a richer understanding of the cognitive processes underlying human learning and exploration. While we found evidence that time pressure reduces directed exploration—consistent with directed exploration being a controlled and reasoned process—we did not predict uncertainty avoidance under time pressure. Together with the finding that relative uncertainty slowed down reaction times and dampened evidence accumulation, our results suggest that time pressure does not eliminate the ability to track uncertainty. Rather, it alters attitudes towards it, from seeking out uncertainty to avoiding it. Future studies should therefore investigate the conditions that cause uncertainty-seeking or uncertainty-avoidance and test whether uncertainty-avoidance is a deliberate behavior (Schulz, Klenske, Bramley, & Speekenbrink, 2017).

Acknowledgments

CMW is supported by the International Max Planck Research School on Adapting Behavior in a Fundamentally Uncertain World; ES is supported by the Harvard Data Science Initiative

References

Annis, J., Miller, B. J., & Palmeri, T. J. (2017). Bayesian inference with Stan: A tutorial on adding custom distributions. *Behavior Research Methods*, *49*, 863–886.

Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, *47*, 235–256.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7–15.

Boldt, A., Blundell, C., & De Martino, B. (2017). Confidence modulates exploration and exploitation in value-based learning. *bioRxiv*, 236026.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.

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.

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature neuroscience*, *3*(11s), 1218–1223.

Gershman, S. J. (2015). A unifying probabilistic view of associative learning. *PLoS Computational Biology*, *11*, e1004567.

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, *173*, 34–42.

Gershman, S. J. (in press). Uncertainty and exploration. *Decision*.

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society. Series B (Methodological)*, 148–177.

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2017). Bridgesampling: an r package for estimating normalizing constants. *arXiv preprint arXiv:1710.08162*.

Hoffman, M. D., & Gelman, A. (2014). The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–1623.

Knox, W. B., Otto, A. R., Stone, P., & Love, B. (2012). The nature of belief-directed exploratory choice in human decision-making. *Frontiers in Psychology*, *2*, 398.

Nursimulu, A. D., & Bossaerts, P. (2013). Risk and reward preferences under time pressure. *Review of Finance*, *18*, 999–1022.

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, *2*, e55.

Schulz, E., & Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Current Opinion in Neurobiology*, *55*, 7–14.

Schulz, E., Klenske, E., Bramley, N., & Speekenbrink, M. (2017). Strategic exploration in human adaptive control. *bioRxiv*, 110486.

Schulz, E., Konstantinidis, E., & Speekenbrink, M. (2015). Learning and decisions in contextual multi-armed bandit tasks. In *Thirty-Seventh Annual Conference of the Cognitive Science Society*.

Schulz, E., Wu, C. M., Huys, Q. J., Krause, A., & Speekenbrink, M. (2018). Generalization and search in risky environments. *Cognitive science*, *42*, 2592–2620.

Schulz, E., Wu, C. M., Ruggeri, A., & Meder, B. (2018). Searching for rewards like a child means less generalization and more directed exploration. *bioRxiv preprint*.

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

Stojic, H., Schulz, E., Analytis, P. P., & Speekenbrink, M. (2018). It's new, but is it good? How generalization and uncertainty guide the exploration of novel options. *PsyArXiv preprint*.

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, *143*, 155-164.

Wu, C. M., Schulz, E., Garvert, M. M., Meder, B., & Schuck, N. W. (2018). Connecting conceptual and spatial search via a model of generalization. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th annual conference of the cognitive science society* (pp. 1183–1188). Austin, TX: Cognitive Science Society.

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2018). Generalization guides human exploration in vast decision spaces. *Nature Human Behaviour*, *2*, 915—924.

Yechiam, E., Busemeyer, J. R., Stout, J. C., & Bechara, A. (2005). Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. *Psychological Science*, *16*, 973–978.

Yu, A. J., & Dayan, P. (2003). Expected and unexpected uncertainty: ACh and NE in the neocortex. In *Advances in Neural Information Processing Systems* (pp. 173–180).