Neuroscience

Human Exploration Strategically Balances Approaching and Avoiding Uncertainty

Reviewed Preprint

Published from the original preprint after peer review and assessment by eLife.

About eLife's process

Reviewed preprint version 1
February 15, 2024 (this version)

Posted to preprint server
December 8, 2023

Sent for peer review
December 8, 2023

Yaniv Abir ✉, Michael N. Shadlen, Daphna Shohamy ✉

Department of Psychology, Columbia University, New York, NY, USA • Zuckerman Mind Brain Behavior Institute, and Kavli Institute for Brain Science, Columbia University, New York, NY, USA • Department of Neuroscience and Howard Hughes Medical Institute, Columbia University, New York, NY, USA

Abstract

A central purpose of exploration is to reduce goal-relevant uncertainty. Consequentially, individuals often explore by focusing on areas of uncertainty in the environment. However, people sometimes adopt the opposite strategy, one of avoiding uncertainty. How are the conflicting tendencies to approach and avoid uncertainty reconciled in human exploration? We hypothesized that the balance between avoiding and approaching uncertainty can be understood by considering capacity constraints. Accordingly, people are expected to approach uncertainty in most cases, but to avoid it when overall uncertainty is highest. To test this, we developed a new task and used modeling to compare human choices to a range of plausible policies. The task required participants to learn the statistics of a simulated environment by active exploration. On each trial, participants chose to explore a better-known or lesser-known option. Participants generally chose to approach uncertainty, however, when overall uncertainty about the choice options was highest, they instead avoided uncertainty and chose to sample better-known objects. This strategy was associated with faster decisions and, despite reducing the rate of observed information, it did not impair learning. We suggest that balancing approaching and avoiding uncertainty reduces the cognitive costs of exploration in a resource-rational manner.

eLife assessment

This study presents a **valuable** investigation of how people approach and avoid uncertainty, with a particular focus on the effects of overall uncertainty. They find that individuals approach uncertainty to a point, but when uncertainty is particularly high, they avoid it. The results are interpreted under a cognitive cost-resource rational framework. The methods are **convincing**, using appropriate and current methodologies, but more details on analyses and placing the work more fully in the context of the existing literature would make the contribution more significant.

Introduction

The purpose of exploration is to reduce uncertainty about the aspects of one's environment that are goal relevant or otherwise important. Yet, devising an optimal strategy to reduce uncertainty is known to be very difficult (Cohen et al., 2007; Schulz and Gershman, 2019; Sutton and Barto, 2018), especially for agents with limited memory and processing capacities. A heuristic strategy that is often efficient for exploration is focusing on the parts of the environment that one is most uncertain about. This principle of approaching uncertainty has been applied in a range of fields, including statistics (MacKay, 1992; Sebastiani and Wynn, 2000), artificial intelligence (Badia et al., 2020; Bellemare et al., 2016; Pathak et al., 2017; Raposo et al., 2021), and cognitive theories of human exploration (Schulz and Gershman, 2019; Schwartenbeck et al., 2019). Indeed, humans have been shown to approach uncertainty when learning about rewards in the environment through trial and error (Schulz and Gershman, 2019; Speekenbrink and Konstantinidis, 2015; Wilson et al., 2014; Wu et al., 2022).

However, there are also many examples of uncertainty avoidance in the decision making of humans and animals. Uncertainty avoidance has been documented in situations where resolving uncertainty may reveal negative outcomes and news (Ahmadlou et al., 2021; Botta et al., 2020; Eilam and Golani, 1989; Glickman and Sroges, 1966; Gordon et al., 2014; Gigerenzer and Garcia-Retamero, 2017; Golman et al., 2017), or may make overcoming a conflict in motivation more difficult (Carrillo and Mariotti, 2000; Golman et al., 2017). When the goal is to maximize immediate rewards, choosing the most rewarding option often entails avoiding more uncertain options (Trudel et al., 2020; Wilson et al., 2014).

How are the two conflicting tendencies to approach and avoid uncertainty reconciled when exploring? To answer this question, we must address gaps in the literature about exploration at two levels of analysis. At the computational level, it is unclear what might compel individuals to avoid uncertainty instead of approaching it, bar holding goals other than attaining knowledge. Indeed, avoiding uncertainty reduces the rate of information intake, and so might result in poorer learning. At the algorithmic level, we lack an understanding of how individuals compute uncertainty to make exploratory choices. Tallying uncertainty in an exact manner is complicated and often intractable. Several candidate algorithms for approximating the computation of uncertainty have been suggested (Schulz and Gershman, 2019), but evidence as to their use by humans is still preliminary.

It is the complexity of choosing based on uncertainty, set against the limited processing and memory capacities that are inherent to human cognition, that motivated our hypotheses regarding both the algorithmic and computational questions. First, we charted a hypothesis space of plausible algorithms for computing uncertainty and making exploratory choices (Schulz and Gershman, 2019), starting with the optimal but complex, and ending with simple approximations. Second, we hypothesized that the complexity of choosing what to explore, even when using approximate algorithms, is the key factor explaining why and when individuals might avoid uncertainty in exploration. Adhering to the goal of approaching uncertainty may well be an efficient policy for an agent with unlimited cognitive resources. Since humans have finite memory systems, inference bandwidth, and time, it stands to reason that they would try to conserve these resources by regulating their exploration (Lieder and Griffiths, 2020), possibly by selectively avoiding uncertainty. Following this insight, we examined exploratory choices as a function of two factors affecting the difficulty of making an exploratory choice: participants' overall uncertainty about choice options (Schulz and Gershman, 2019), and forgetting.

We developed a task requiring participants to make multiple exploratory choices, incrementally building knowledge in the service of a distant goal (**Figure 1**). Importantly, participants were given reward feedback only at the end of a round and not after every trial, allowing us to focus on choices made to accumulate knowledge, rather than choices driven by the need to exploit available rewards. Seeking ecological validity, we designed a task that posed a challenging exploration problem for participants, requiring that they infer and remember the values of multiple latent parameters from repeated experience (Hartley, 2022; Lieder and Griffiths, 2020). The task could nonetheless be captured by a few mathematical expressions, allowing for the derivation of the optimal exploration policy. This optimal policy served as a basis for a quantitative analysis of participants' choices and reaction times with the aim of identifying the algorithm driving their exploratory choices (Anderson, 1990; Chater and Oaksford, 1999; Waskom et al., 2019).

Results

194 participants from a pre-registered (Abir et al., 2021) sample were recruited to complete up to 22 rounds of the exploration task over four online sessions. The task simulated a room with four tables, with two decks of cards on each table (**Figure 1** a-b). If a card was flipped, it was revealed to be, for example, either orange or blue (each round used a different pair of colors). The proportion of orange vs. blue cards, π differed between the two decks on each table. Participants' goal was to learn *sgn(*$\pi_1 - \pi_2$ or which deck had more orange (blue) cards on each table. We will denote this term, which serves as the learning desideratum for participants, as θ.

The task begins with an exploration phase, followed by a test phase. On each trial of the exploration phase participants chose which of two tables to explore, and then revealed one card from a deck on that table (**Figure 1 b**). Participants were instructed that the exploration phase would be followed by a test phase after a random number of trials (drawn from a geometric distribution to discourage pre-planning, **Figure 1c**). They were further instructed that one of the colors would be designated as rewarding at the beginning of the test phase. During the test phase, participants were asked to indicate which deck had more of the rewarding color on each table (**Figure 1b**). They also rated their confidence in the choice. For every correct test-phase choice they received $0.25. Crucially, they received no reward during exploration. Participants' only incentive during the exploration phase was to maximize their confidence about the value of θ.

Three Hypothetical Strategies Derived by Rational Analysis

To explain how participants chose between tables in the exploration phase, we first asked how an optimal agent might solve the problem of choosing which table to explore on each trial of the task. We limited our consideration to strategies that optimize learning only for the next trial, since a globally optimal strategy is intractable for this task (Schulz and Gershman, 2019; Sutton and Barto, 2018). We started by deriving the optimal strategy and progressively simplified it to generate two additional strategies. While they differ in the level of complexity they assume, all three strategies direct an agent using them to approach the option they are more uncertain about.

The optimal strategy, given by the expression at the top of **Figure 1d**, is choosing the table affording maximal expected information gain (EIG; Gureckis and Markant (2012); MacKay (1992); Yang et al. (2016)). EIG is the difference between the uncertainty in the value of the learning desider-atum, θ, given observed cards $x_{0:\,t}$ and the expected uncertainty after observing the next card on trial t +1. In other words, EIG is the amount of uncertainty resolvable on the next trial.

Computing the second term in the EI expression requires averaging over future unseen outcomes, which may be beyond the ability of participants. As an alternative, they might avoid computing this term by simply choosing the table they were more uncertain about at the moment of making the choice (**Figure 1d**, second tier; Schulz and Gershman, 2019). While this strategy has intuitive appeal, computing uncertainties may still be too complicated for human participants. An even simpler heuristic is given on the third tier of **Figure 1** d: choosing the table with the least prior exposure (Auer, 2002; Schulz and Gershman, 2019), measured as the number of already

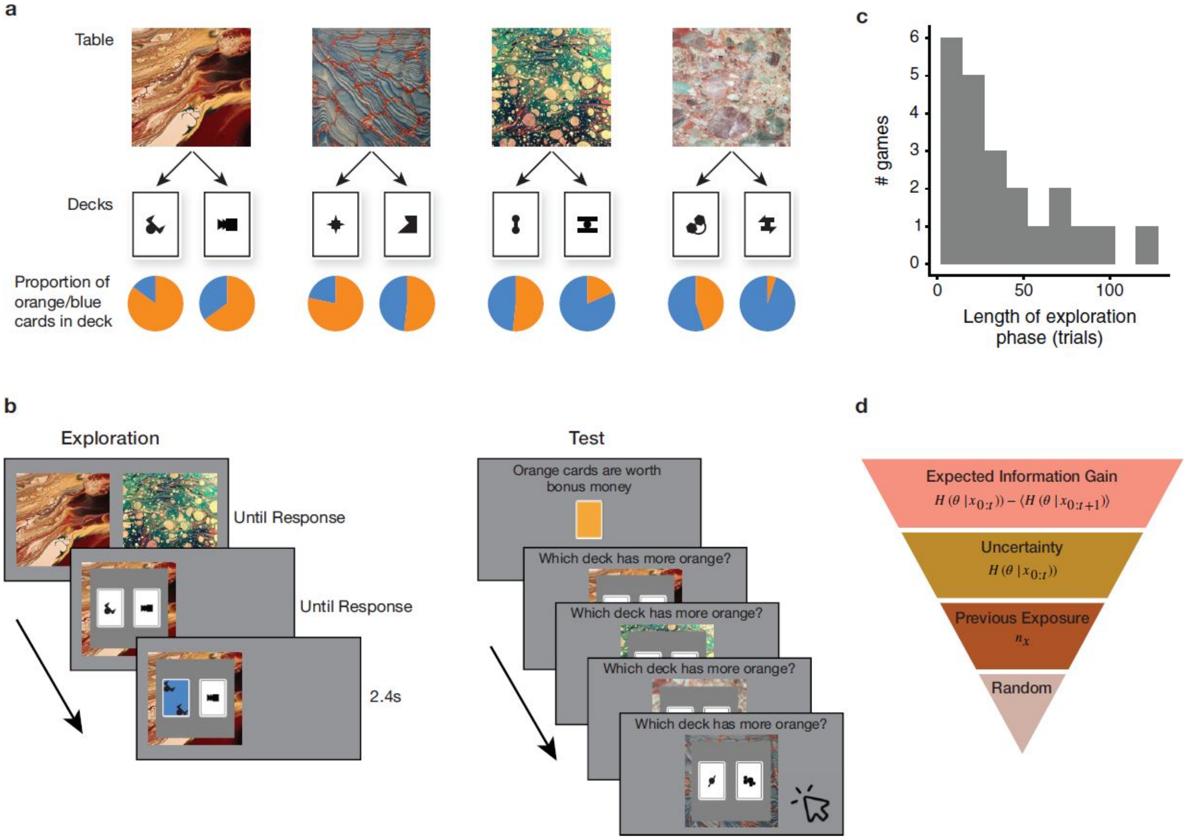

a
Table
Decks
Proportion of orange/blue cards in deck
c
# games
6
5
4
3
2
1
0
0
50
100
Length of exploration phase (trials)
b
Exploration
Until Response
Until Response
2.4s
Test
Orange cards are worth bonus money
Which deck has more orange?
Which deck has more orange?
Which deck has more orange?
Which deck has more orange?
d
Expected Information Gain
$H(\theta \mid x_{0:t}) - \langle H(\theta \mid x_{0:t+1}) \rangle$
Uncertainty
$H(\theta \mid x_{0:t})$
Previous Exposure
n_x
Random

Figure 1.

Examining exploration strategy in relation to uncertainty in an incremental learning task.

a, Structure of the task. Participants explored four tables, each containing two decks with different proportions of blue/orange cards. The goal was to learn the difference in proportions of the decks on each table. **b**, The two phases of the task - exploration and test. On a single exploration trial (left), participants chose between two tables, and then sampled a card from one of the decks on that table, observing its color. After a random number of exploration trials, participants were tested on their knowledge (right). A color was designated as rewarding, and participants then chose the deck with the highest proportion of the rewarding color on each table. They were rewarded for correct test-phase choices, and received no reward during exploration. **c**, Histogram of round lengths. Participants played 22 rounds. The length of exploration in each round followed a shifted geometric distribution, such that the test was equally likely to occur following any trial after the first 10. **d**, We considered a hierarchy of strategies for choosing which table to explore. The normatively prescribed strategy is to choose the table affording maximal expected information gain. This is the table for which the next card is expected to maximally decrease uncertainty (measured as entropy *H*) about the value of the goal-relevant latent parameter *θ,* given observations thus far *x.* A simpler strategy is to choose the table with the maximum uncertainty, as it does not necessitate computing an expectation over the next observation. An even simpler heuristic is to equate previous exposure and choose the table with the least previous observations n_x. Even though these three strategies vary considerably in complexity, they are all uncertainty-approaching on average. Lastly, people may be random explorers.

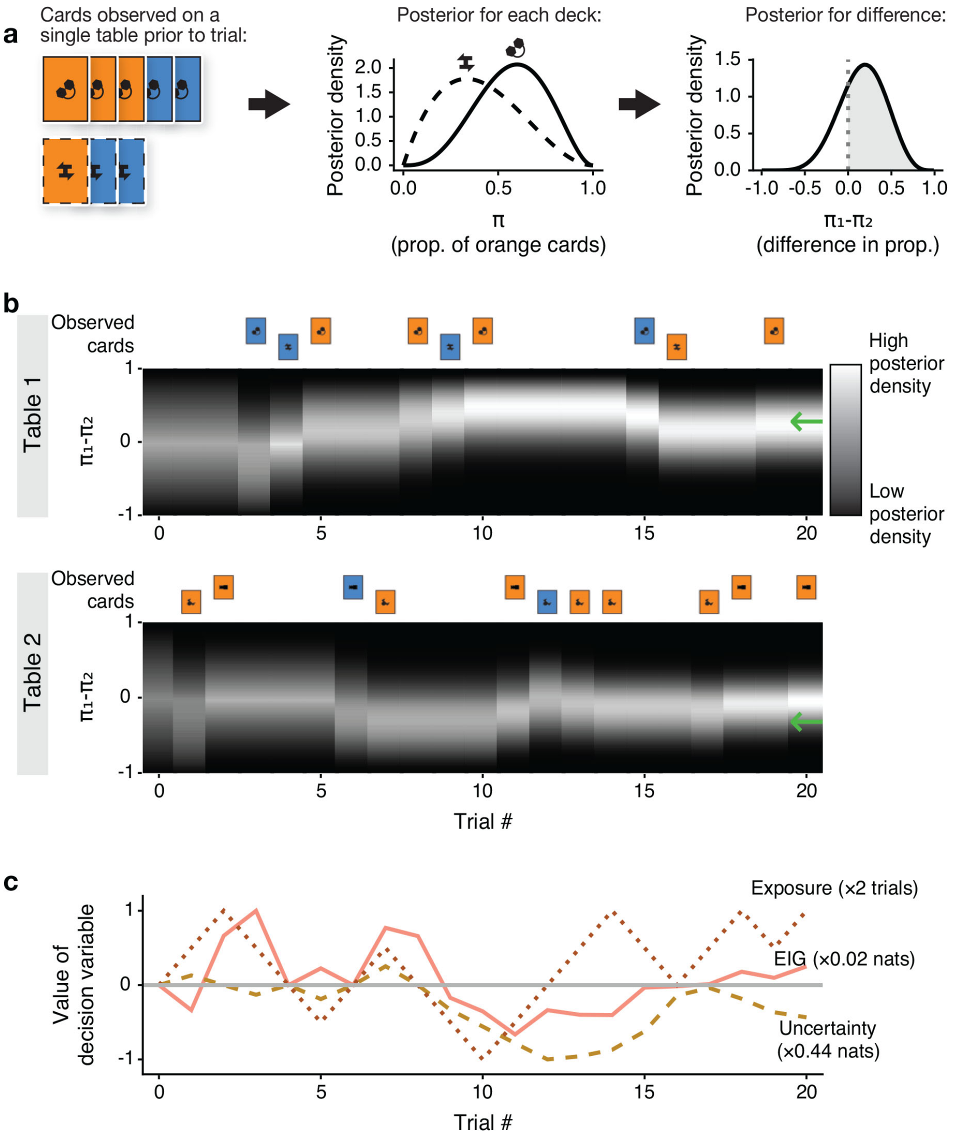


Figure 2.

Hypothetical strategies make differing predictions for exploratory choice behavior.

We computed the three quantities hypothesized to drive exploratory choices using a Bayesian observer model. To illustrate this process, we plot the derivation of Bayesian belief on a single trial. (**a**) and across multiple trials (**b**, **c**). For visualization, we use a simplified version with two tables only. **a** depicts the Bayesian observer's belief about a single table on a single trial. Given a sequence of previously observed cards (left), the Bayesian observer forms posterior beliefs about the proportion of orange cards in each deck (center). These beliefs are expressed as Beta distributions. From these, it is possible to derive a belief about the difference in the proportion of orange cards between the two decks $\pi_1 - \pi_2$ (right). The probability that $\pi_1 <$ π_2 is given by the proportional size of the area marked in gray (0.74 in this example). **b** Depicts the same process over a series of20trials. The observed card sequence for each table is presented at the top of each panel. The matching belief state about $\pi_1 - \pi_2$ is plotted below it as an evolving posterior density in white (high) and black (low). The green arrows mark the true value of $\pi_1 - \pi_2$ for that round. As the round progresses, belief converges towards the true value, and becomes more certain. **c**, The three choice strategies prescribe different table choices on most trials. The difference between **table 1** and **table 2** in each of the three quantities (EIG, uncertainty and exposure) is plotted for each trial. This difference is the hypothesized decision variable for choosing between **tables 1** and **2**. A positive value indicates a preference for exploring table 1, and a negative value a preference for **table 2**. The three variables are normalized to facilitate visual comparison.

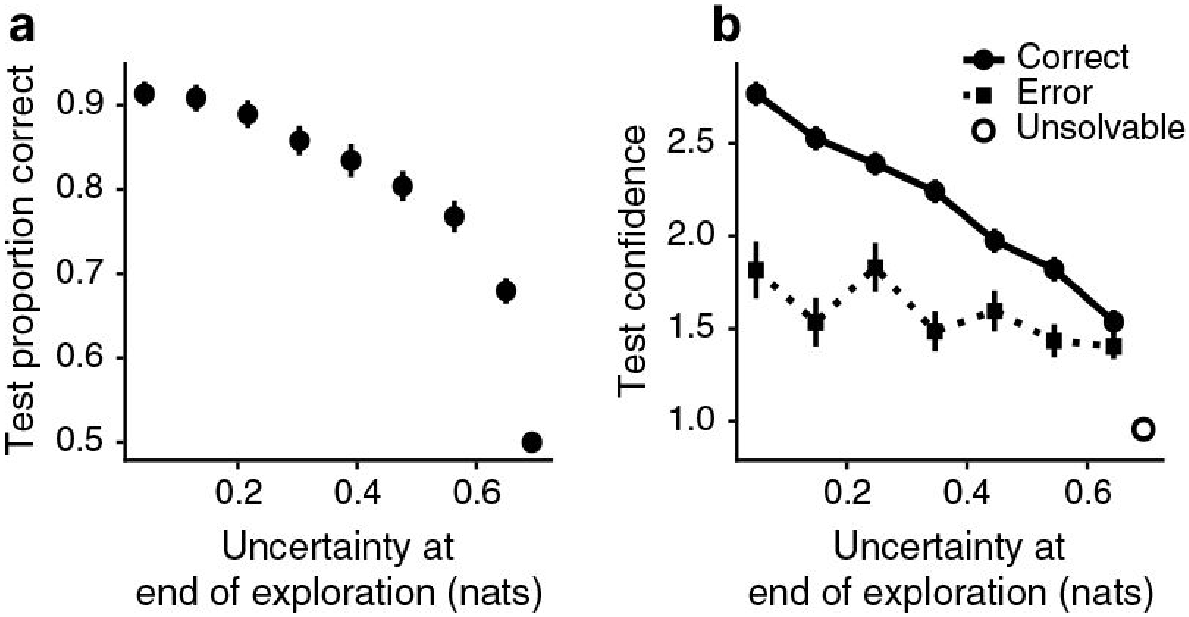


Figure 3.

The Bayesian observer model is validated by participants' accuracy and confidence on the test phase.

a, Participants were accurate when an exploration phase ended with low uncertainty, and performed at chance level when the phase ended with high uncertainty. **b**, Participant's confidence on correct choices fell with rising uncertainty. Confidence on error trials did not depend as much on Bayesian observer uncertainty. When a test question was unsolvable because no evidence was observed on each deck during exploration, participants had very low confidence. Data presented as mean values ±1 SE, n=194 participants.

Figure 3—figure supplement 1. Matching results in the preliminary sample.

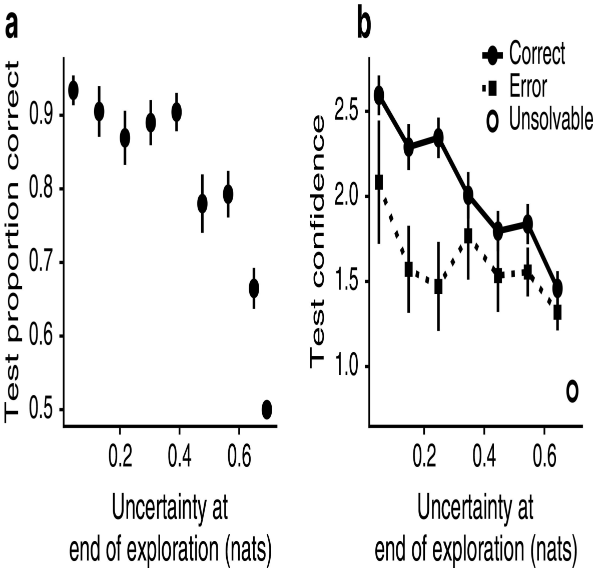


Figure 3—figure supplement 1.

Reproducing the analysis using the preliminary sample: The Bayesian observer model is validated by participants' accuracy and confidence on the test phase.

a, Participants were accurate when an exploration phase ended with low uncertainty, and performed at chance level when the phase ended with high uncertainty. **b**, Participant's confidence on correct choices fell with rising uncertainty. Confidence on errors did not depend as much on Bayesian observer uncertainty. Data presented as mean values ±1 SE, n=62 participants. Nats are the units of entropy, a mathematically convenient measure of uncertainty.

observed cards n_x. Since on average additional observations result in lower uncertainty, this strategy is an approximate way to approach the more uncertain table. Finally, participants might explore at random, rather than in a directed manner (Daw et al., 2006; Schulz and Gershman, 2019; Wilson et al., 2014).

Test Phase Performance Validates Observation Model

To relate the three hypothesized strategies to participants' behavior, we assumed a model of participants' beliefs about the goal-relevant parameter θ and the mechanism by which they updated these beliefs. We used a Bayesian observer model which forms beliefs about θ based on the actual card sequence each participant observed, and updates these beliefs according to Bayes' rule (**Figure 2**). On its own, the Bayesian observer does not predict participants' exploration choices, but only models the process of inference from observation.

Before evaluating the hypothesized exploration strategies, we sought to validate the assumptions of the Bayesian observer model. To this end, we related the predictions of the Bayesian observer model to participants' choices during the test phase. We predicted that test accuracy should be greater when the Bayesian observer model had low uncertainty about θ at the end of the learning phase. The data supported this prediction (**Figure 3**). Using a multilevel logistic regression model, we confirmed that test accuracy was strongly related to the Bayesian observer's uncertainty b=-5.59, 95% posterior interval (PI)=[-6.25,-4.95] (all effect sizes given in original units, full model and coefficients reported in Appendix 3—**Table 1**). Participants' reports of confidence after making a correct choice also followed the Bayesian observer's uncertainty b=-4.04, 95% PI=[- 4.50,-3.56]. After committing errors, participants' reported confidence was lower overall b=-1.09, 95% PI=[-1.27,-0.92], and considerably less dependent on Bayesian observer uncertainty, interaction b=-3.10, 95% PI=[-3.76,-2.46] (**Figure 3b**, Appendix 3—**Table 2**).

Uncertainty is the Best Predictor of Exploratory Choice

To evaluate the three exploration strategies, we tested whether participants' exploration-phase choices could be predicted from the difference between the two tables that were presented as choice options in each of the hypothesized quantities. We fit the data with a multilevel logistic regression model for each strategy (Appendix 3—**Tables 3**-**5**). In a formal comparison of the three models we found that uncertainty was the best predictor of exploratory choices, as indicated by a reliably better prediction metric (**Figure 4**). Accordingly, the difference in uncertainty for the table presented on the right versus the table presented on the left (Δ-uncertainty) predicts participants' choices. Δ-EIG provides a poorer fit to choices, and Δ-exposure is anti-correlated with choice, in contradiction of the exposure hypothesis. We confirmed that our analysis approach can recover the true model generating a simulated dataset(**Figure 4—figure Supplement 1**). Furthermore, simulations showed that uncertainty is a sufficient predictor of choice. Simulated datasets generated by uncertainty-driven agents recreated the entire set of qualitative and quantitative results (**Figure 4—figure Supplement 2**). The simulations demonstrate that the surprising negative correlation between choice and Δ-exposure is an epiphenomenon of uncertainty-based exploration.

Participants Systematically Change Their Exploration Strategy According to Overall Uncertainty

We next asked whet her participants' strategy of exploring by approaching uncertainty is modulated by the state of their knowledge when making an exploratory choice. Specifically, we examined how participants' overall uncertainty about the two options they could choose to explore on a given trial changed the way the explored (**Figure5a**). Since table choice options were presented at random, participants sometimes had to choose between tables they already knew a lot about, and sometimes between tables they were very uncertain about. When overall

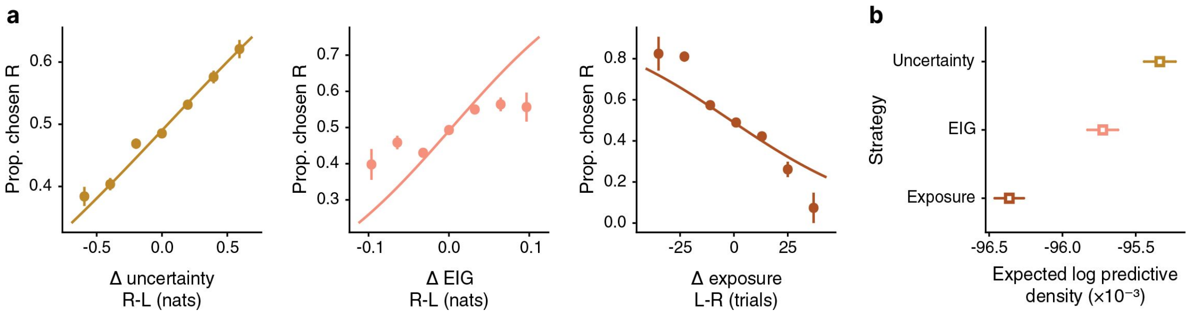


Figure 4.

Uncertainty is the best predictor of choice.

a, On each plot the difference in the hypothesized quantity between the two tables presented on each trial is plotted against actual choices of the table presented on the right. For each plot, the relevant hypothesis predicts a positive smooth curve. - uncertainty, plotted on the left, matches this prediction better than Δ (center). The relationship between Δ-exposure (right) and choice is negative, rather than the hypothesized positive correlation. **b** Quantitative model comparison confirms this observation. Out of the three hypothesized strategies uncertainty has the highest approximate expected log predictive density (using PSIS LOO; see Methods). Data presented as mean values ±1SE, n=194 participants.

Figure 4—figure supplement 1. Fitting simulated data successfully recovers the underlying strategy.

Figure 4—figure supplement 2. Uncertainty is a sufficient predictor of choice.

Figure 4—figure supplement 3. Matching results in the preliminary sample.

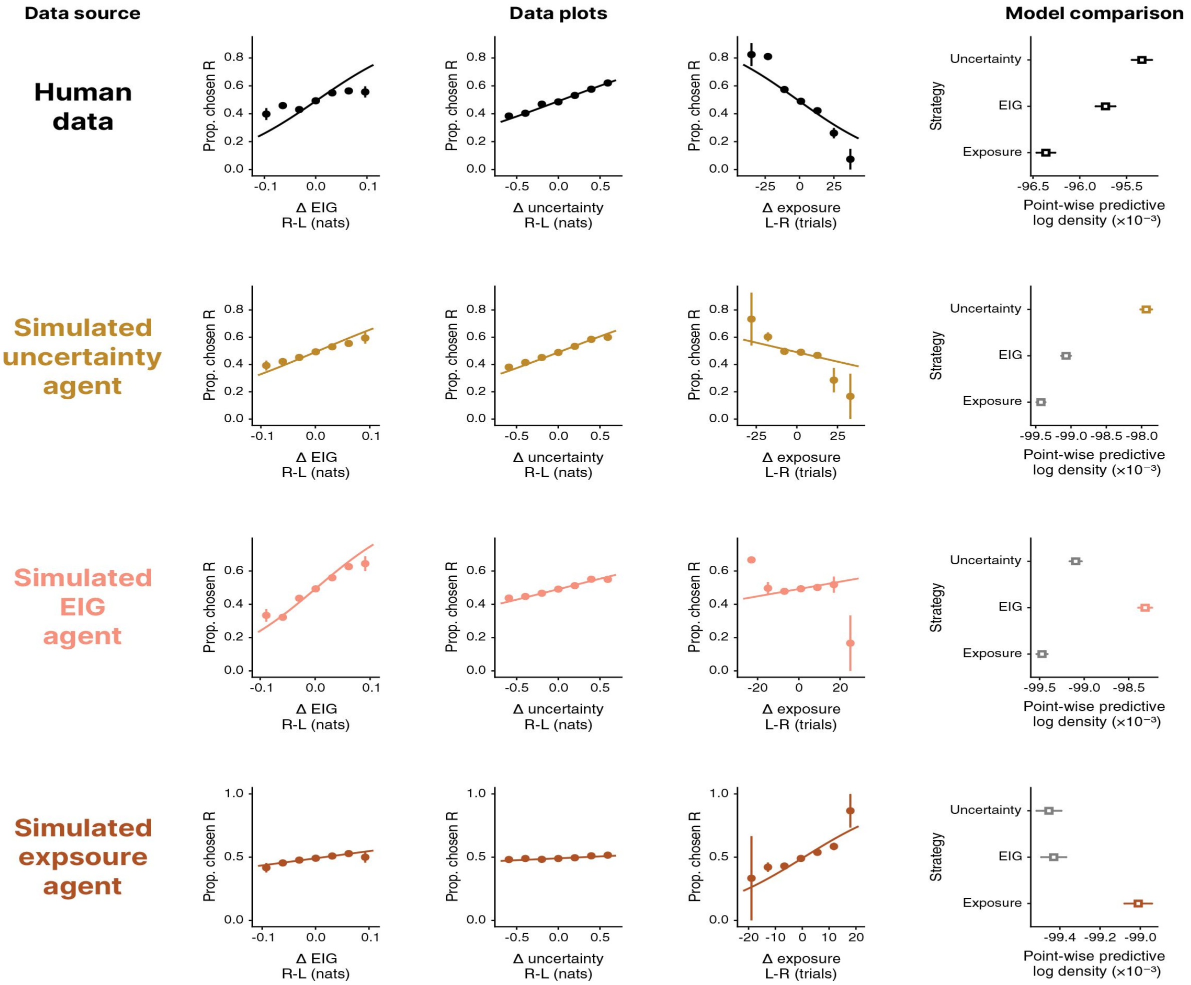


Figure 4—figure supplement 1.

Our analysis approach successfully recovers the strategy used by simulated agents.

We compared the actual data (top) to datasets generated by artificial agents. Each simulated dataset comprised a group of agents operating according to one of the hypothesized strategies. We fixed the effect size for each strategy in the simulations to the effect size we observed for uncertainty in the actual data. Each agent matched a single participant in the true dataset, choosing to observe cards from the same decks presented to the participant. Each agent chose the table on the right or the table on the left on each trial, with the probability of choosing the table on the right given by *f (a + b × Δx),* where *f* is the logistic function, *b* is the degree to which the agent's choices are dependent on *Ax,* standing for the relevant decision variable, and *a* is a general bias towards rightward or leftward choices. Coefficients *a* and *b* were extracted per participant from the uncertainty model described in **Figure 4**. For the sake of this analysis, we assumed the agents choose a random deck on the table of their choice. Here, the simulated data for each group of agents was plotted against each of the three decision variables and fit with the same models we used on the actual dataset (center). We tested whether our procedure for qualitative and quantitative model comparison used in **Figure 4** is potent at recovering the true strategy generating the data. For easy comparison, the actual data is re-plotted on the first row. For each of the three simulated strategies, we observe successful recovery: the decision variable matching the true strategy shows the strongest positive correlation with choice (center), and the correct strategy is indicated as best fitting the data (right). Furthermore, a negative correlation between behavior and Δ-exposure, as observed in the true data, is only evident in the uncertainty-based group of agents (second row). Data plotted as means ±1SE, n=194 participants/agents.

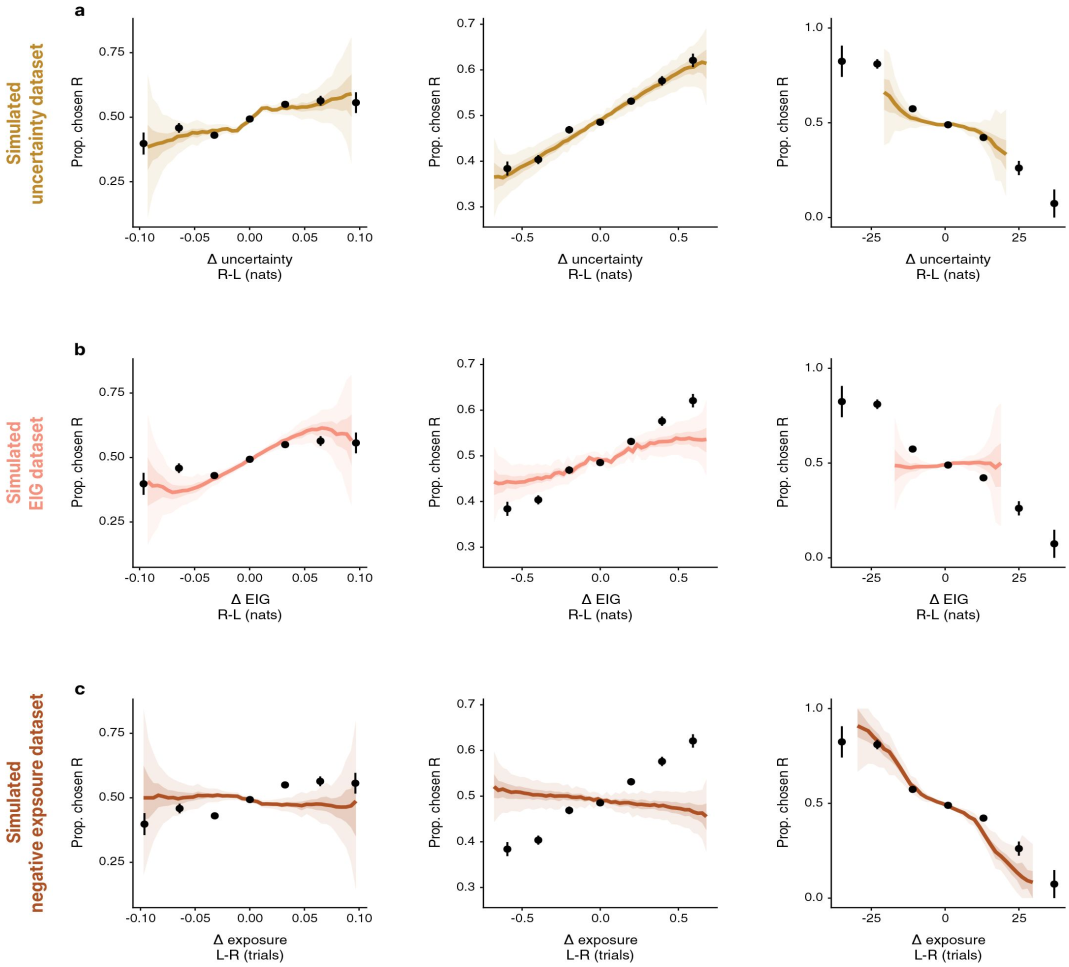


Figure 4—figure supplement 2.

Simulations confirm that uncertainty is a sufficient predictor of choice.

We further confirmed our conclusion that uncertainty is the best predictor of participants' choices, by plotting the posterior predictive distribution for each of the models predicting choice from a hypothesized strategy. We simulated 500 datasets for each of the three models, and plotted the distribution of the simulated data (green lines with 50% and 95% posterior interval bands, n=500 iterations) against the observed dataset (means ±1SE plotted in black, n=194 participants). The simulation procedure was similar to that used in **Figure 4—figure Supplement 1**, with the exception that coefficients were extracted from the posterior distribution of each model fitted to the actual data. We expect that the posterior predictive distribution for each model would capture the relationship between the relevant decision variable and choice well. The extent to which the posterior predictive distribution can recreate the association with the other two decision variables is a test of model fit. **a**, The posterior predictive distribution for the EIG model does not match the observed data well: it does not reproduce the strong slope for Δ-uncertainty, nor the negative correlation with Δ-exposure. **b**, The posterior predictive distribution for uncertainty captures the data very well, matching the particular shape of the correspondence between choices and Δ-EIG, and the negative correlation between choices and. Δ-exposure. **c**, The posterior predictive distribution of exposure does not match observe data well: it fails to recreate the positive correlations between choice and EIG and uncertainty.

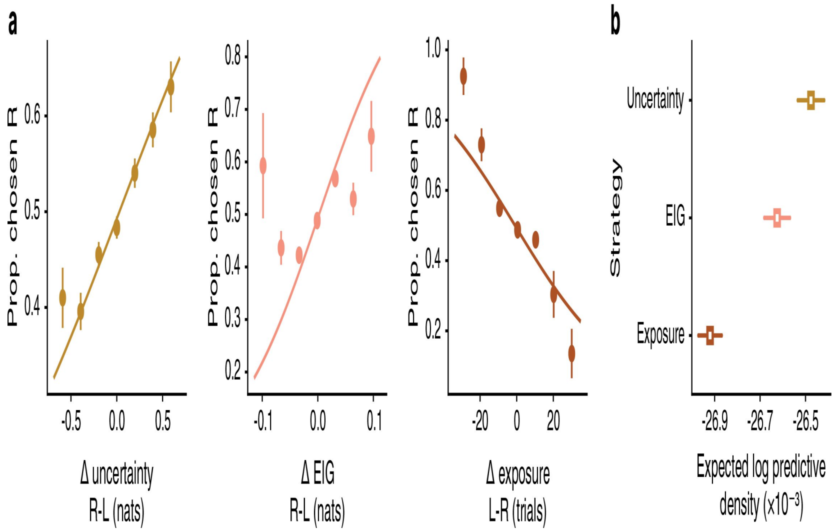


Figure 4—figure supplement 3.

Reproducing the analysis in Figure 4 using the preliminary sample: Uncertainty is the best predictor of choice.

a, On each plot the difference in the hypothesized quantity between the two tables presented on each trial is plotted against actual choices of the table presented on the right. For each plot, the relevant hypothesis predicts a positive smooth curve. Δ-uncertainty, plotted on the left, matches this prediction better than Δ-EIG (center). The relationship between Δ-exposure (right) and choice is negative, rather than the hypothesized positive correlation. **b** Quantitative model comparison confirms this observation. Out of the three hypothesized strategies uncertainty has the highest approximate expected log predictive density (PSIS LOO; see Methods). Data presented as mean values ±1 SE, n=62 participants.

uncertainty was high, the choice between tables had to be made with very little evidence. Note that from a normative perspective, choice should follow the difference in uncertainty between options and shouldn't be influenced by overall uncertainty.

We found a systematic deviation in exploration strategy in relation to overall uncertainty. When overall uncertainty for the two choice options was below a certain threshold, participants chose the more uncertain table, as expected. However, when overall uncertainty was above the threshold, they chose the less uncertain table, thereby slowing the rate of information intake (**Figure 5b, c**).

We validated this observation using a multilevel piecewise-regression model, allowing for the influence of Δ-uncertainty on choice to differ below and above a fitted threshold of overall uncertainty. We observed a positive relationship between Δ-uncertainty and choice below the threshold b=0.97, 95% PI=[0.83, 1.11], but above the threshold we found that the influence of Δ-uncertainty on choice became strongly negative (interaction b=-4.3e+02, 95% PI=[-5.4e+02,-3.4e+02]]). The threshold was estimated to be 1.28 nats of overall uncertainty (95% PI=[1.27, 1.29]; Appendix 3— **Table 6**), leaving 21.58% of trials in the high overall uncertainty range (95% PI=[20.12,24.45]). This bias in exploration cannot be viewed merely as a noisier version of optimal performance. Rather, it constitutes a systematic modulation of exploration strategy on about a fifth of the trials.

Costs and Benefits of Strategically Avoiding Uncertainty

What motivates participants to systematically avoid learning about more uncertain objects? By the standards of an ideal agent, uncertainty avoidance is clearly suboptimal, as it reduces the rate of observed information, and thus the potential capacity to learn. We hypothesized that the limited processing and memory capacities that are inherent to human cognition and set it apart from the optimal agent are the reason for uncertainty avoidance. To test this hypothesis, we conduct a cost benefit analysis of uncertainty avoidance in the following sections. We ask whether uncertainty avoidance is associated with costs to learning, and whether it affords any benefits in managing cognitive effort.

Uncertainty Avoidance is not Associated with Learning Deficits

Since efficient learning is the purpose of exploration, we asked how the tendencies to approach uncertainty and avoid it when overall uncertainty is high affect learning as reflected in performance attest. If approaching uncertainty is the only rational exploration policy, then participants who tend to approach uncertainty to a greater degree should learn more and perform better at test, while participants with a strong tendency to avoid uncertainty should learn less and perform worse at test, since they are choosing to forgo valuable information as they explore.

To test these predictions we examined individual differences in exploration strategy in relation to test performance. We found that participants' baseline tendency to approach uncertainty predicted better performance at test b=2.96, 95% PI=[2.67, 3.25] (**Figure 6b**; Appendix 3—**Table 7**). In contrast, we found no evidence that participants with a strong tendency to avoid uncertainty performed worse at test. Indeed, a stronger tendency to avoid uncertainty when overall uncertainty is high was associated with a small improvement in test performance b=1.18, 95% PI=[0.80, 1.58] (**Figure 6b**; Appendix 3—**Table 8**). Thus, modulating exploration according to overall uncertainty was not maladaptive, resulting in no decrement to learning. This result suggests that the rate of information intake is not the limiting factor for the efficiency of exploration and learning.

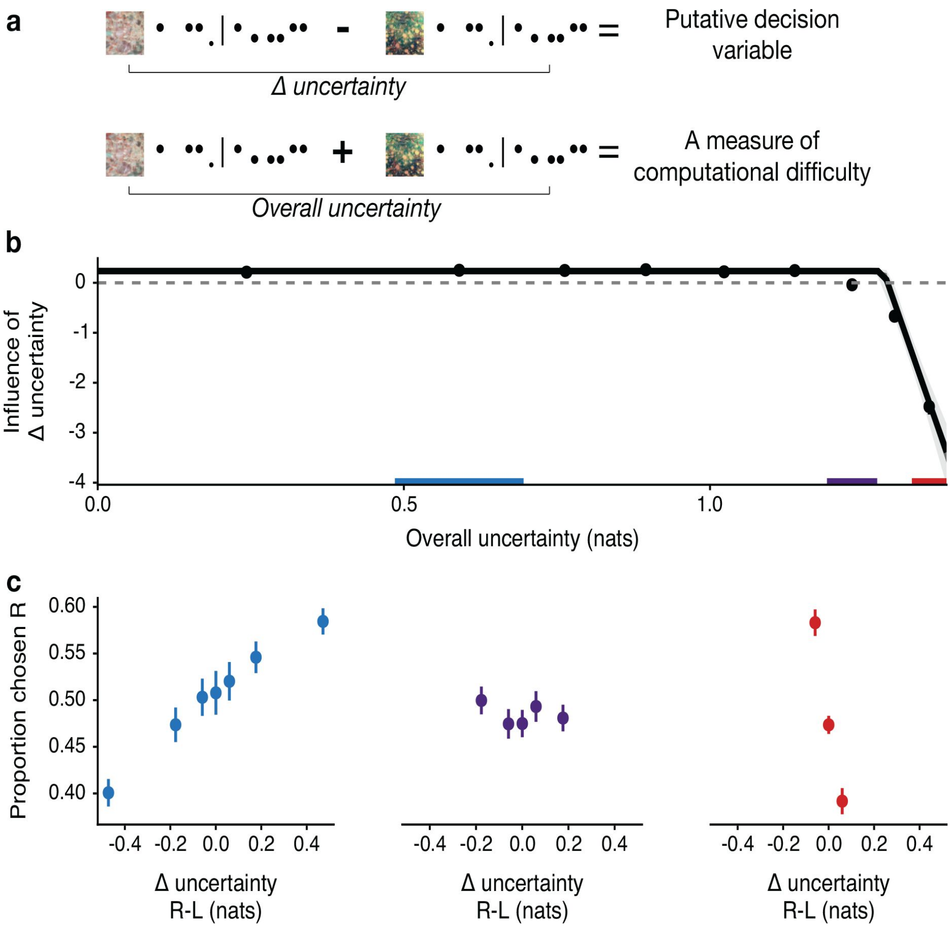


Figure 5.

Participants approach vs. avoid Δ-uncertainty as a function of overall uncertainty.

a, While the Δ-uncertainty is the decision variable identified above, overall uncertainty, defined as the sum of uncertainty for both tables, is a measure of decision difficulty. **b**, The influence of Δ-uncertainty on choice differed markedly below and above a threshold of overall uncertainty. Below an estimated threshold of overall uncertainty, Δ-uncertainty had a significant positive effect on choice. Above this threshold of overall uncertainty, the influence of Δ-uncertainty became strongly negative. Points denote mean posterior estimate from regression models fitted to binned data, error bars mark 50% PI. The solid line depicts the prediction from a piecewise regression model capturing the non-linear relationship and estimating the threshold, with darker ribbon marking50%PI and light ribbon marking95% PI. Data from three regions of overall uncertainty marked in color are plotted in **c**. For low overall uncertainty (blue) participants tend to choose the table they are more uncertain about, as normatively prescribed. But that relationship is broken for medium levels of overall uncertainty (purple). For high overall uncertainty (red), participants strongly prefer to choose the table they are less uncertain about, thereby slowing down the rate of information intake. Data plotted as mean ±SE, n=194 participants.

Figure 5—figure supplement 1. Matching results in the preliminary sample.

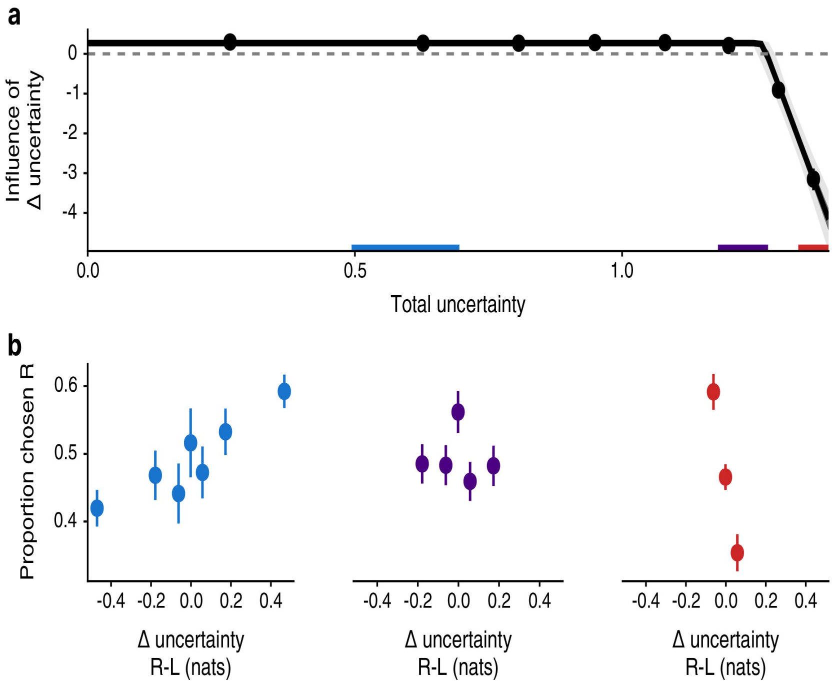


Figure 5—figure supplement 1.

Reproducing the analysis using the preliminary sample: Participants approach vs. avoid Δ-uncertainty as a function of overall uncertainty.

a, The influence of Δ-uncertainty on choice differed markedly below and above a threshold of overall uncertainty. Below a certain estimated threshold of overall uncertainty, Δ-uncertainty had a significant positive effect on choice. Above this threshold of overall uncertainty, the influence of Δ-uncertainty decreased significantly. Points denote mean posterior estimate from regression models fitted to binned data, error bars mark 50% PI. The solid line depicts the prediction from a piecewise regression model capturing the non-linear relationship and estimating the threshold, with the darker ribbon marking 50% PI and the light ribbon marking 95% PI. Data from three regions of overall uncertainty marked in color are plotted in b. For low overall uncertainty (blue) participants tend to choose the table they are more uncertain about, as normatively prescribed. But that relationship is broken for medium levels of overall uncertainty (purple). For high overall uncertainty (red), participants strongly prefer to choose the table they are less uncertain about, thereby slowing down the rate of information intake. Data plotted as mean ±SE, n=62 participants.

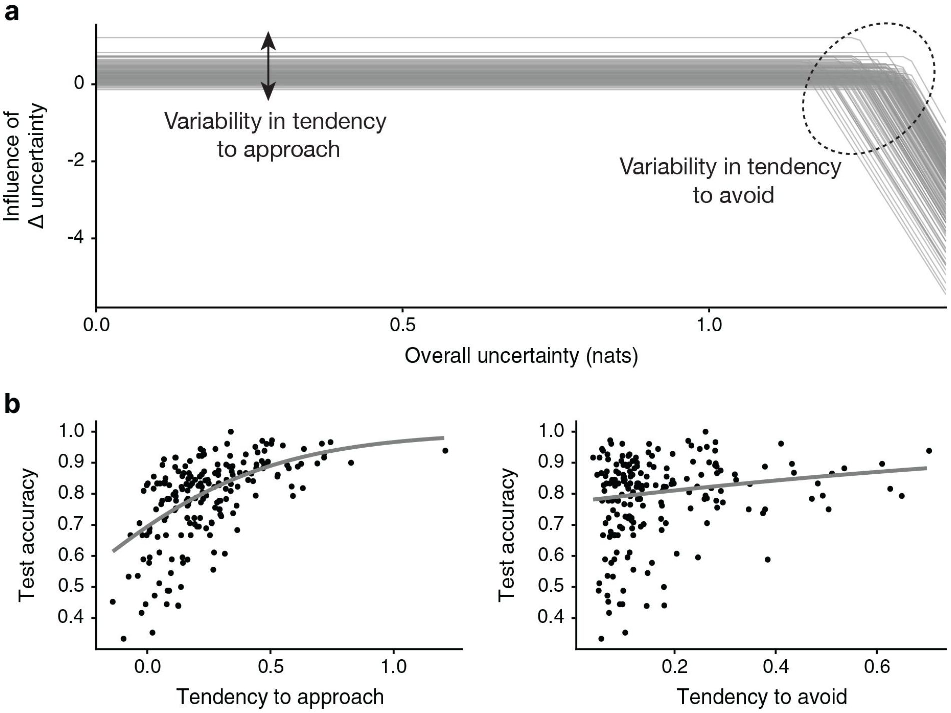


Figure 6.

Approaching uncertainty benefits learning while avoiding uncertainty does not hurt it.

a, We observe substantial individual differences in strategy. Replotting **Figure 5** e for each individual reveals differences in the baseline tendency to approach uncertainty, and differences in the interaction with overall uncertainty, which captures uncertainty avoidance when overall uncertainty is high. **b**, Associations between test performance and the parameters describing approaching and avoiding uncertainty. The baseline tendency to approach uncertainty (left) is strongly associated with performance at test, such that participants who are unable to approach uncertainty also perform poorly at test. There is a weak and positive correlation between test performance and the tendency to avoid uncertainty when overall uncertainty is high (right), indicating that uncertainty avoidance does not hinder learning. Uncertainty avoidance is quantified based on the individual lines plotted in panel **a** as the triangular area charted by the piecewise regression line.

Figure 6—figure supplement 1. Matching results in the preliminary sample.

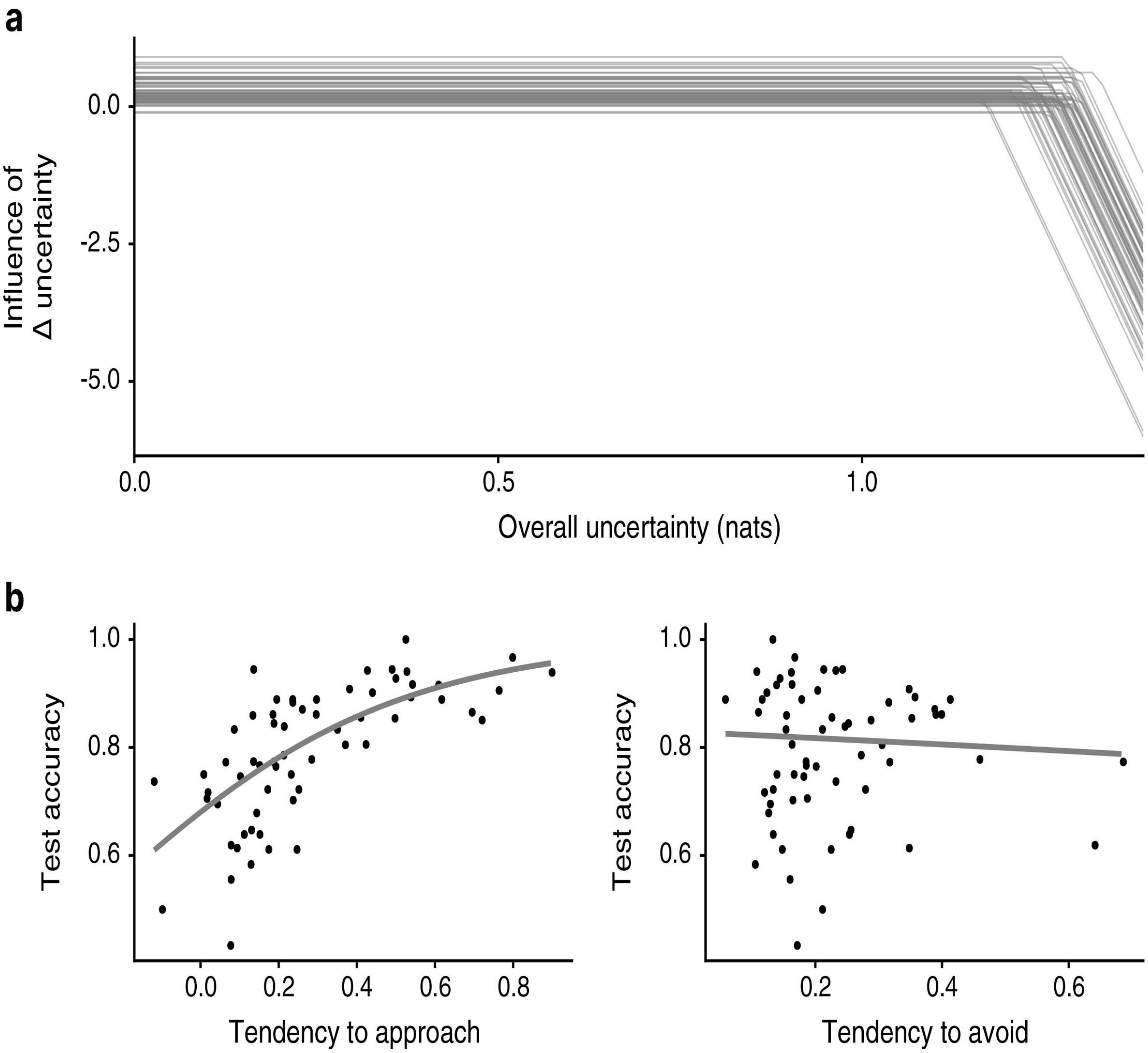


Figure 6—figure supplement 1.

Reproducing the analysis using the preliminary sample: Approaching uncertainty benefits learning while avoiding uncertainty does not hurt it.

a, We observe substantial individual differences in strategy. Replotting **Figure 5e** for each individual reveals differences in the baseline tendency to approach uncertainty, and differences in the interaction with overall uncertainty, which captures uncertainty avoidance when overall uncertainty is high. **b**, Associations between test performance and the parameters describing approaching and avoiding uncertainty. The baseline tendency to approach uncertainty (left) is strongly associated with performance at test, such that participants who are unable to approach uncertainty also perform poorly at test. There is a weak and positive correlation between test performance and the tendency to avoid uncertainty when overall uncertainty is high (right), indicating that uncertainty avoidance does not hinder learning. Uncertainty avoidance is quantified based on the individual lines plotted in panel **a** as the triangular area charted by the piecewise regression line.

Strategic Exploration Involves Costly Deliberation

To understand the costs involved in exploration, we asked whether making exploratory choices in this task involves prolonged deliberation. If that is the case, and exploratory choices are guided by ΔΔ-uncertainty, we reasoned that decisions should require longer deliberation when the absolute value of Δ-uncertainty is small (Palmer et al., 2005; Shushruth et al., 2022). To test this prediction, we fit the data with a generative model of choice and RTs. We used a sequential sampling model, which explains decisions as the outcome of a process of sequential sampling that stops when the accumulation of evidence satisfies a bound. This model explains RTs as jointly influenced by participant's efficacy in deliberating about Δ-uncertainty, and their tendency to deliberate longer vs. make quick responses (Ratcliff and McKoon, 2008; Shadlen and Kiani, 2013; Shadlen and Shohamy, 2016). One prediction of sequential sampling theory is that greater deliberation efficacy should be manifested as as greater dependence of RT on absolute Δ-uncertainty (Palmer et al., 2005).

We found that RTs indeed varied in relation to the absolute value of Δ-uncertainty as expected b=0.69, 95% PI=[0.58,0.78] (Appendix 3—Table 9). Crucially, a strong dependence of RT on the absolute value of Δ-uncertainty predicted better performance at test b=0.81, 95% PI=[0.58,1.07]. We further found that participants who tended to deliberate longer for the sake of accuracy also tended to perform better at test b=1.46, 95% PI=[0.58,2.34] (**Figure 7c**, Appendix 3—**Table 10**). In summary, participants who were better at deliberating about uncertainty during exploration, and who deliberated for longer, performed better at test. Thus, making good exploratory choices that lead to efficient learning involves prolonged deliberation.

Deliberation is Reduced by Choice Repetition

We have shown that when overall uncertainty is high participants avoid uncertainty rather than approach it, and that they do not pay a learning cost as a result. It remains to be shown that participants' alternative strategy is able to shorten the time spent deliberating. Unfortunately, we could not test for such a benefit by directly comparing RTs as a function of overall uncertainty, as overall uncertainty is related to the difficulty of making an exploratory choice. With a single independent variable, deconfounding the effect of difficulty from the strategies used to ameliorate it is impossible. Fortunately, we could take advantage of a conceptually-related but independent tendency we observed in our dataset to examine the benefits of reduced deliberation times.

As in many learning tasks, participants in our task tended to repeat their previous choice (Wu etal.,2022),a tendency that was in dependent of Δ-uncertainty or overall uncertainty. We observed that participants generally preferred to re-choose the table they had last chosen (**Figure 8b**). We corroborated this with a multilevel regression model controlling for the effects of Δ-uncertainty and overall uncertainty b=0.50, 95% PI=[0.42,0.59] (Appendix 3—**Table 11**). Crucially, the tendency to repeat choices was also reflected in RTs, which for repeat choices were less related to Δ-uncertainty (b=-0.32, 95% PI=[-0.43,-0.22]). We also found that participants tended to make repeat choices more quickly rather than deliberate longer (b=-0.05, 95% PI=[-0.05,-0.04]; **Figure 8c**, Appendix 3— **Table 12**).

As in other aspects of exploration strategy, we observed considerable individual differences in the tendency to repeat previous choices. These differences were associated with the uncertainty based aspects of exploration discussed above (**Figure 8d**). Participants with a general tendency to repeat choices show stronger uncertainty avoidance when overall uncertainty is high, indicating that these two conceptually related strategies also co-occur in the population r=-0.60, 95% PI=[-0.74,-0.43] (Appendix 3—**Table 11**). Furthermore, the tendency to repeat previous choices is associated with better test performance, logistic regression b=0.09, 95% PI=[0.07,0.11] (Appendix 3—**Table 13**). The tendency to repeat is also correlated with a stronger baseline tendency to approach uncertainty r=0.32, 95% PI=[0.17,0.46] (Appendix 3—**Table 11**), which was shown above to be correlated with test performance. Thus, while from a normative point of view repeating the previous choice appears to be a context-insensitive heuristic, in practice participants who use this strategy do not learn any worse.

Forgetting as a Conceptual Control

Explaining participants' deviation from the optimal exploration strategy as rational is interesting only to the extent that rationality is not a forgone conclusion. Is the alternative hypothesis of a failure in decision making also a-priori plausible?. We turned to forgetting as a second source of difficulty in our task and a conceptual control condition. Due to the random presentation of choice options, there was variability in the number of trials passed since either of the presented tables was last explored. We assumed that choosing between tables that had not been explored fora long time is more difficult than between tables for which evidence has been recently observed. Indeed, we found that RTs were longer with a larger lag, indicating greater difficulty of making a choice (log normal regression b=0.02, 95% PI=[0.02, 0.03]; **Figure 9a**, Appendix 3—**Table 14**). Furthermore, we observed that exploration choices on trials with a greater lag depended less on Δ-uncertainty b=- 0.08,95% PI=[-0.11,-0.04], and that the tendency to repeat the last chosen table on these trials was also diminished b=-0.13, 95% PI=[-0.15, -0.11] (**Figure 9b**, Appendix 3—**Table 15**). Finally, on trials with a large lag the difference in RTs between making a repeat and a switch choice disappeared, interaction b=0.02, 95% PI=[0.02,0.03] (**Figure 9a**, Appendix 3—**Table 14**). These patterns suggest that prior evidence is forgotten with increasing lag and that as a consequence exploration becomes more random. Hence, in contrast to the systematic effect of overall uncertainty, forgetting results in a failure to make principled exploratory choices.

Discussion

We examined the cognitive computations behind exploratory choices using a paradigm that encourages incremental learning in the service of a distant goal. We found that uncertainty played an important role in guiding participants' choices about how to sample their environment for learning. In general, participants chose to learn more about the options they were more uncertain about. However, when overall uncertainty was especially high, participants instead avoided the more uncertain options and sampled the options they already knew more about. In addition, we found that participants tended to repeat previous choices. Together, this pattern suggests that participants systematically balance approaching and avoiding uncertainty while exploring.

Examining individual differences in exploration and learning revealed the costs and benefits of avoiding uncertainty when exploring. We found that strategically avoiding uncertainty is not associated with a detriment to learning, even though it slows down the rate of information intake. We also found an association between the length of deliberation and learning efficiency. Participants who deliberated longer also learned better, and deliberation time could be shortened by repeating previous choices. Based on these results, we conclude that balancing approaching and avoiding uncertainty is a way to manage cognitive resources by regulating deliberation costs. In this sense, our results serve as an example of how human cognition is adapted to the inherent constraints of the human mind, consistent with the resource rationality framework (Lieder and Griffiths, 2020).

While the literature on exploration is expansive, the paradigm presented here extends it in important ways. Researchers of reinforcement learning have previously examined how exploration manifests when agents learn incrementally about their environment. Crucially, this literature has focused on cases where reward can be gained on each trial (Brown et al., 2022; Cohen et al., 2007; Daw et al., 2006; Schulz and Gershman, 2019; Song et al., 2019; Tversky and Edwards, 1966; Wilson et al., 2014; Wu et al., 2022). In contrast, our task was designed to remove the impetus to exploit current knowledge immediately, a motivation that

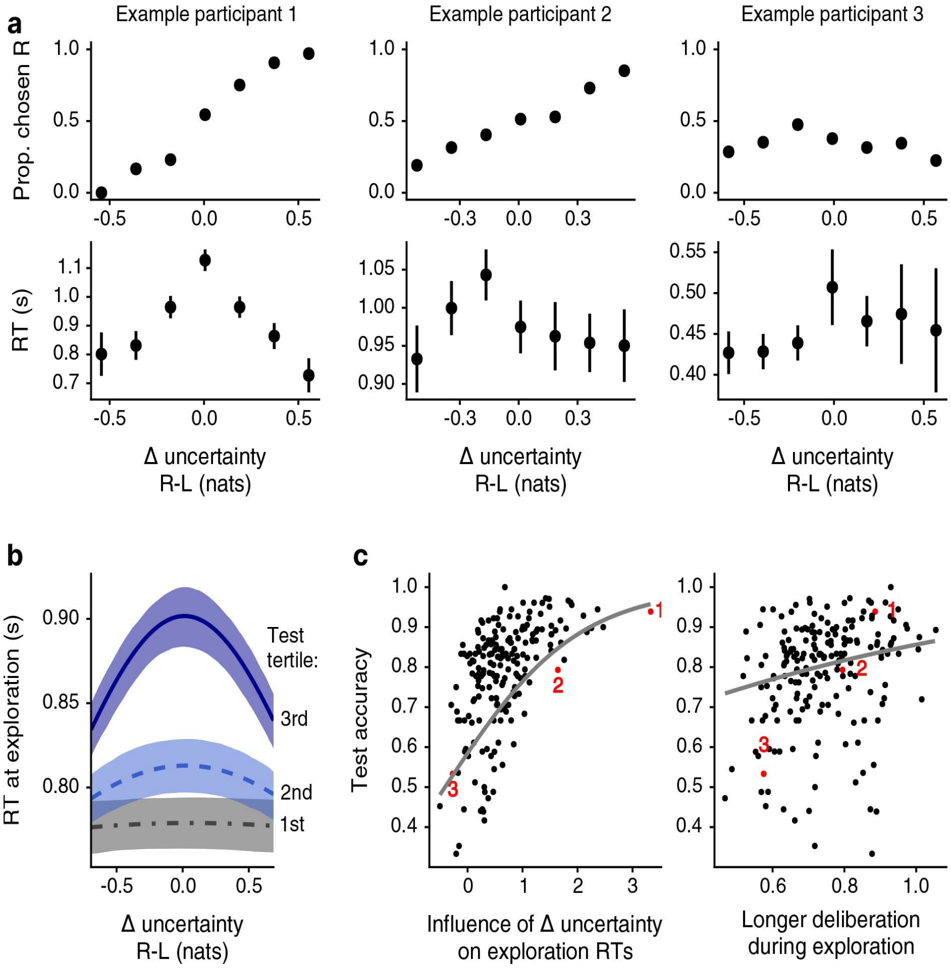


Figure 7.

Individuals who spend time deliberation during exploration make strategic choices and learn well.

Participants varied not only in the pattern of their choices, but also in their RTs. **a**, Data from three example participants. The relationship of choice and RTs with Δ-uncertainty weakens from left to right. Data plotted as mean ±SE. **b**, These individual differences were captured by a sequential sampling model, explaining choices and RTs as the interaction between participant's efficacy of deliberating about Δ-uncertainty and their tendency to deliberate longer vs. make quick responses. Plotting model predictions, we observe a u-shaped dependence of RTs on Δ-uncertainty for participants whose performance at test was in the top accuracy tertile. This characteristic u-shape is indicative of decisions made by prolonged deliberation. This relationship is weaker for participants in the bottom two test accuracy tertiles. Such participants also exhibit shorter RTs overall. Lines mark mean predictions from a sequential sampling model fit by tertiles for visualization, ribbons denote 50% PIs. **c**, Correlating the sequential sampling model parameters with test performance confirms these observations. Participants with a stronger dependence of RT on Δ–uncertainty perform better at test, as do participants who deliberate longer for the sake of accuracy. Example participants from **a** are marked in red. Lines are mean predictions from a logistic regression model.

Figure 7—figure supplement 1. Matching results in the preliminary sample.

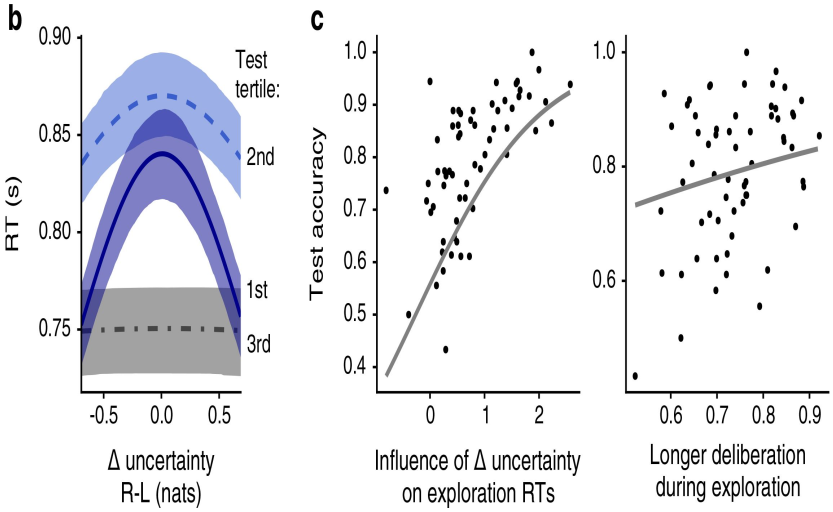


Figure 7—figure supplement 1.

Reproducing the analysis using the preliminary sample: Individuals who spend time deliberation during exploration make strategic choices and learn well.

Participants varied not only in the pattern of their choices, but also in their RTs. **a**, Data from three example participants. The relationship of choice and RTs with Δ-uncertainty weakens from left to right. Data plotted as mean ±SE. **b**, These individual differences were captured by a sequential sampling model, explaining choices and RTs as the interaction between participant’s efficacy of deliberating about Δ-uncertainty and their tendency to deliberate longer vs. make quick responses. Plotting model predictions, we observe a u-shaped dependence of RTs on Δ-uncertainty for participants whose performance attest was in the top accuracy tertile. This characteristic u-shape is indicative of decisions made by prolonged deliberation. This relationship is weaker for participants in the bottom two test accuracy tertiles. Such participants also exhibit shorter RTs overall. Lines mark mean predictions from a sequential sampling model fit by tertiles for visualization, ribbons denote 50% PIs. **c**, Correlating the sequential sampling model parameters with test performance confirms these observations. Participants with a stronger dependence of RT on Δ-uncertainty perform better at test, as do participants who deliberate longer for the sake of accuracy. Example participants from a are marked in red. Lines are mean predictions from a logistic regression model.

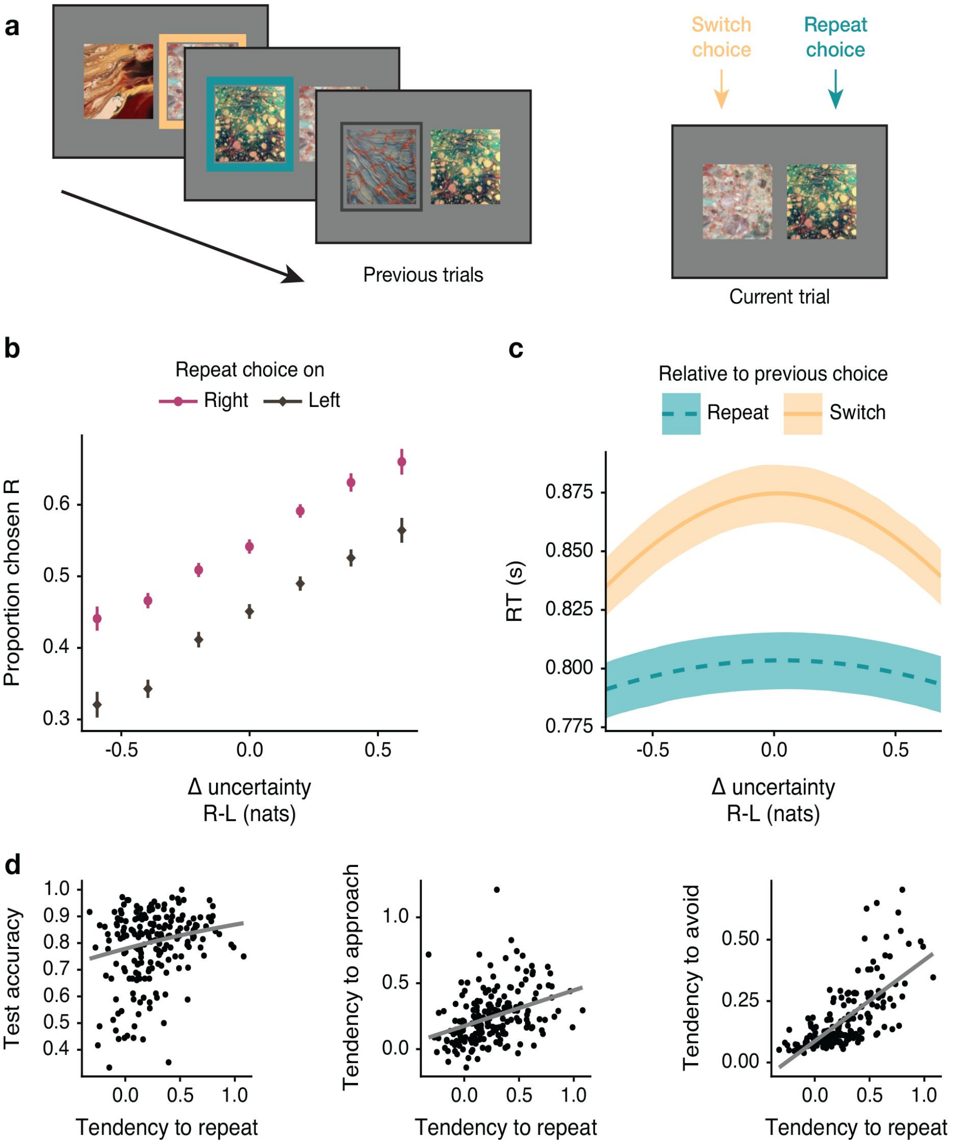


Figure 8.

Participants tend to repeat previous choices instead of deliberating over uncertainty.

a, On a given trial one table has been chosen more recently than the other (frames denote previous choices). In the example the green table had been chosen more recently, hence it is designated the repeat option and the other table the switch option. **b**, Participants tend to choose the table displayed on the right more often when it is the repeat option than when it is the switch option. Data plotted as mean ±SE, n=194 participants. **c**, When choosing a repeat option, participants' RTs are shorter and less dependent on -uncertainty. Lines mark mean predictions from a sequential sampling model, ribbons denote 50% PIs. **d**, Participants who tended to repeat their previous choice also tended to perform better at test (left), were more likely to have a stronger baseline tendency to approach uncertainty (middle), and a stronger tendency to avoid uncertainty when overall uncertainty is high (right). Regression lines are plotted for visualization.
Figure 8—figure supplement 1. Matching results in the preliminary sample.

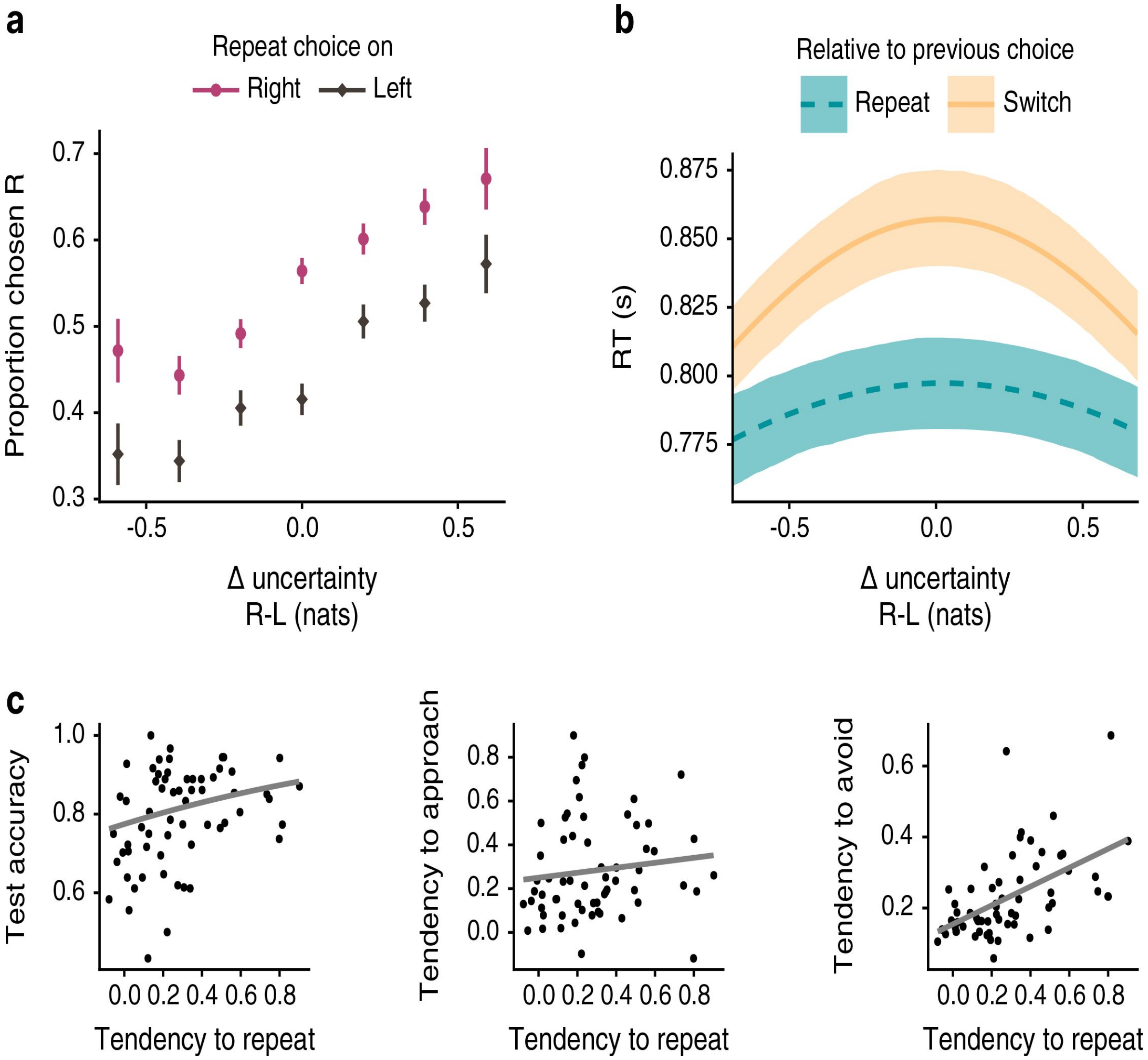


Figure 8—figure supplement 1.

Reproducing the analysis using the preliminary sample: Participants tend to repeat previous choices instead of deliberating over uncertainty.

a, On a given trial one table has been chosen more recently than the other (frames denote previous choices). In the example the green table had been chosen more recently, hence it is designated the repeat option and the other table the switch option. **b**, Participants tend to choose the table displayed on the right more often when it is the repeat option than when it is the switch option. Data plotted as mean ±SE, n=194participants. **c**, When choosing a repeat option, participants' RTs are shorter and less dependent on Δ-uncertainty. Lines mark mean predictions from a sequential sampling model, ribbons denote 50% PIs. **d**, Participants who tended to repeat their previous choice also tended to perform better at test (left), were more likely to have a stronger baseline tendency to approach uncertainty(middle),and a stronger tendency to avoid uncertainty when overall uncertainty is high (right). Regression lines are plotted for visualization.

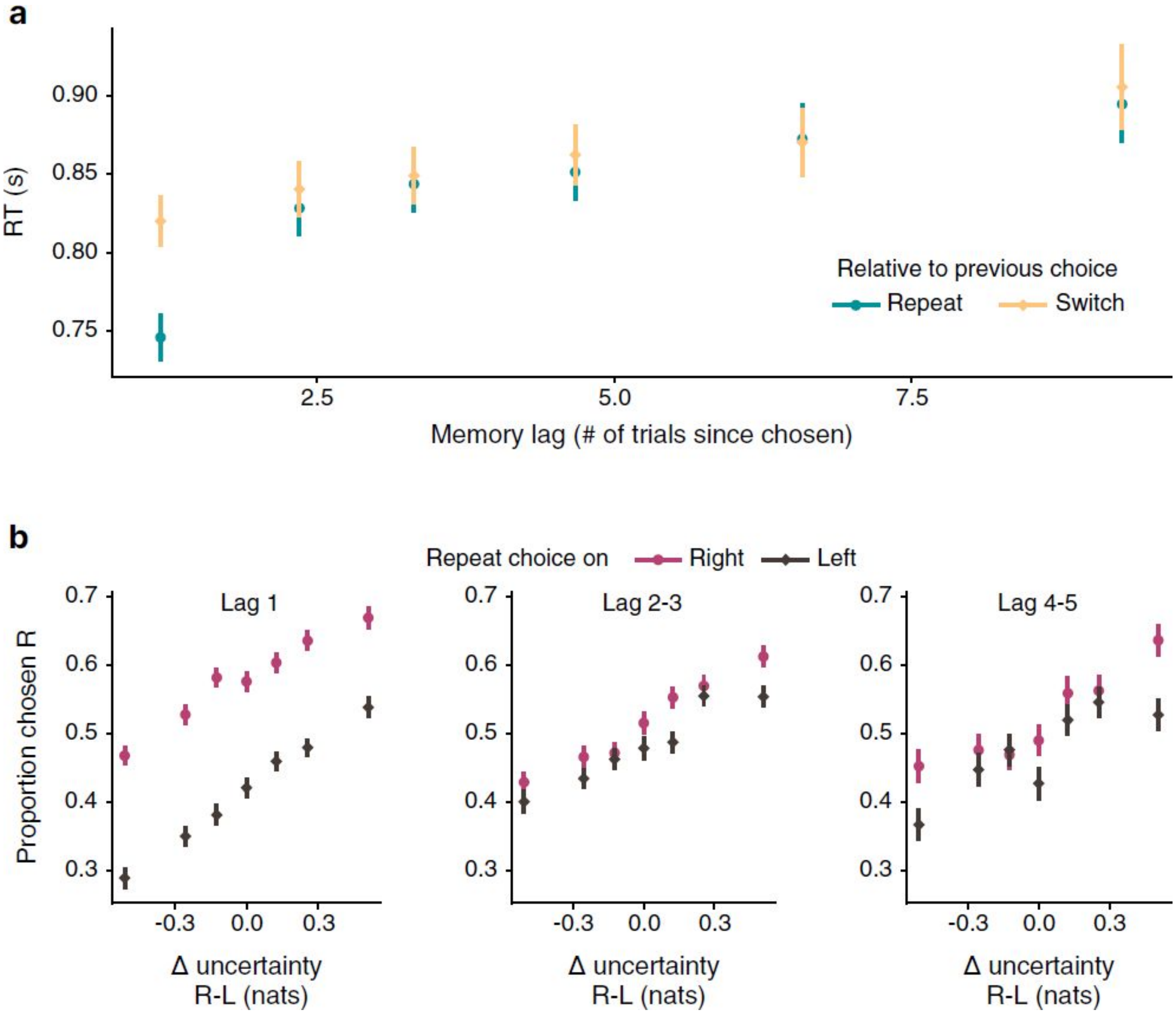


Figure 9.

Forgetting is associated with random choice rather than a systematic bias.

a, Memory lag, defined as trials since last choice, serves as a proxy for forgetting and contributes to the difficulty of making an exploratory choice. RTs rise with memory lag. The RT advantage for repeat choices disappears with higher memory lag. **b**, With higher memory lag choices become less dependent on Δ-uncertainty, as indicated by flatter curves. The tendency to repeat the last choice is also diminished with memory lag. Both effects amount to choice becoming more random due to forgetting. Data plotted as mean ±SE, n=194 participants.

Figure 9—figure supplement 1. Matching results in the preliminary sample.

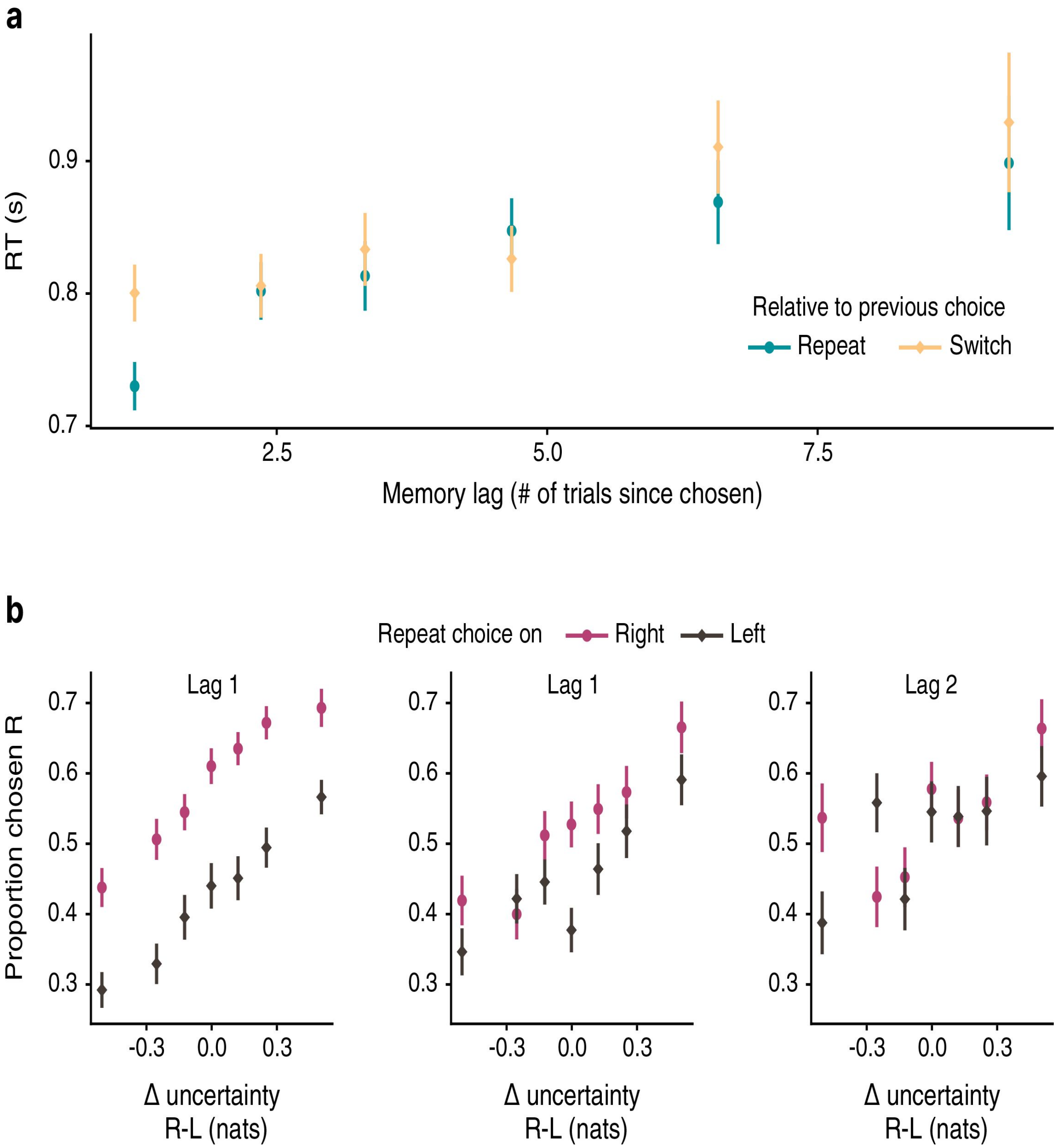


Figure 9—figure supplement 1.

Reproduction the analysis using the preliminary sample.

Forgetting is associated with random choice rather than a systematic bias. **a**, Memory lag, defined as trials since last choice, serves as a proxy for forgetting and contributes to the difficulty of making an exploratory choice. RTs rise with memory lag. The RT advantage for repeat choices disappears with higher memory lag. **b**, With higher memory lag choices become less dependent on Δ-uncertainty, as indicated by flatter curves. The tendency to repeat the last choice is also diminished with memory lag. Both effects amount to choice becoming more random due to forgetting. Data plotted as mean ±SE, n=194 participants.

predominates exploration in tasks with immediate reward. Accordingly, we were able to observe many exploratory choices and had greater experimental power to describe in detail how participants approach uncertainty and when they avoid it instead. Secondly, exploration has been studied in the information search literature (Gureckis and Markant, 2012; Markant and Gureckis, 2014; Oaksford and Chater, 1994; Petitet et al., 2021; Rothe et al., 2018; Ruggeri et al., 2017). In most studies of this field participants make decisions without relying on their memory, as the entire history of learning is displayed to them on screen (cf. related work in active sensing; Yang et al., 2016). This differs from our task, which places heavy demands on memory. Rather than treating capacity limitations as a source of noise and a nuisance to measurement, we find that the rational use of limited resources is central for successful exploration.

Several previous studies of exploration inspired us to design a task with separate exploration and test phases. Using the "observe or bet" paradigm, Tversky and Edwards (1966) examined how participants trade off exploration and exploitation on a trial-by-trial basis. By using a short block of exploration followed by a test, Wilson et al. (2014) achieved a first reliable demonstration of directed exploration in humans. Finally, the expansive literature on the description-experience gap (Wulff et al., 2018) has used a similar paradigm to examine when participants choose to self terminate their exploration, and how that affects their learning. The paradigm presented here extends these approaches, as it is crafted to reveal the strategy driving each exploration choice.

We observed considerable individual differences in exploration strategy, as would be expected in a complex task requiring memory-based learning and inference. In the face of such variability, one may question the prudence of drawing conclusions about the population, since the average might be a poor summary of a plurality of idiosyncratic strategies. However, the strong correlation we observed between individual differences in exploration and test performance mitigates this concern. The correlation suggests that participants who were engaged with this task and able to learn from observation can well be described as exploring by a strategy of combining approaching and avoiding uncertainty. The relationship between test performance and RTs lends additional mechanistic support to this idea.

Our theoretical analysis and experiments leave several questions open. First, overall uncertainty in our task was correlated with the number of cards observed. While our results hold when trial number is added as a covariate to the regression models (see Appendix 3—Table 16), future work orthogonalizing overall uncertainty and time on task would help to fully disentangle the contribution of each factor to uncertainty avoidance.

Another open question is the nature of the limitation driving participants to avoid uncertainty when overall uncertainty is high. This could be due to limitations in committing prior experiences to memory, inferring latent parameters from disparate experiences, retrieving prior knowledge, or estimating the uncertainty of existent knowledge. While the idea that decisions based on high overall uncertainty are more difficult has been raised previously (Schulz and Gershman, 2019; Shafir, 1994), an explanation grounded in cognitive mechanisms is still needed. Accordingly, the mechanism by which uncertainty avoidance ameliorates choice difficulty remains unknown.

One intriguing explanation for the source of difficulty and the way it is managed lies in the distinction between strategies dependent on remembering single experiences and those dependent on the incremental acumulation of knowledge in the form of summary statistics (Collins et al., 2017; Collins and Frank, 2012; Daw et al., 2005; Knowlton et al., 1996; Plonsky et al., 2015; Poldrack et al., 2001). Both strategies could contribute to performance in tasks such as ours. A participant may be encoding prior observations as single instances, or summarizing them into a central tendency with a margin of uncertainty around it. Crucially, each strategy is associated with a different profile of cognitive resource use. Keeping track of individual experiences is much costlier than tracking a single expectation and a confidence interval around it (Daw et al., 2005; Nicholas et al., 2022) and more likely to incur costs when switching between exploring different tables. Prior work suggests individuals use single experiences or summary statistics according to the reliability of each strategy, and the cost of using it (Daw et al., 2005; Nicholas et al., 2022). In our case, summary statistics may be perceived as unreliable when overall uncertainty is high, compelling participants to rely on committing individual experiences to working memory (Bavard et al., 2021; Collins et al., 2017; Daw et al., 2005; Duncan et al., 2019; Poldrack et al., 2001). Furthermore, recent work examining how humans make a series of dependent decisions demonstrates that the tension between remembering single experiences and discarding them in favor of summary statistics is accompanied by a tendency to revisit previous choices instead of switching to new alternatives (Zylberberg, 2021).

The questions we addressed here were partly motivated by the well-established observation that humans and animals often avoid uncertainty in various situations. Two broad categories of explanation for such avoidance have been proposed (Golman et al., 2017). First, individuals avoid resolving uncertainty when it could lead to negative news, for example by avoiding ambiguous prospects when making economic choices (Ellsberg, 1961; Fox and Tversky, 1995). An extension of this idea is dread avoidance (Gigerenzer and Garcia-Retamero, 2017; Golman et al., 2017). One might avoid resolving the uncertainty about a medical diagnosis to avoid the unpleasant affective response to the news, even if the information could be very useful in determining treatment. Relatedly, humans might avoid uncertainty as a by-product of pursuing a goal other than exploration. More uncertain options are often avoided for the sake of choosing immediately rewarding options (Trudel et al., 2020; Wilson et al., 2014) — why try an unknown dish, when your absolute favorite is on the menu? Lastly, uncertainty avoidance may be a strategy for managing conflict between different motivations, or different mechanisms of action selection (Carrillo and Mariotti, 2000; Golman et al., 2017). For example, to maintain their diet, an individual might choose to avoid resolving the uncertainty about what snacks can be found in the office kitchen. Our findings highlight a different kind of strategic uncertainty avoidance. In our tasks there were no negative consequences to learning about the color proportions of card decks, and no conflicting motivations. Rather, we explain participants' tendency to avoid uncertainty in terms of managing their limited cognitive resources.

The idea of a balance between approaching and avoiding uncertainty has conceptual parallels in other literatures. A group of relevant findings concern how animals explore their proximal environment. A classic finding in rats is that when placed in a novel open arena, they alternate between the exploration strategy of walking around the arena (uncertainty approaching) and a strategy of returning to their initial position and pausing there (termed "home base" behavior, which is uncertainty avoiding; Eilam and Golani, 1989). Relatedly, by using computational models to understand how rats use their whiskers to explore near objects, researchers have identified an alternation between uncertainty approaching and avoiding strategies (Gordon et al., 2014). Recent work in mice and primates has uncovered neural circuits driving exploration by framing the problem of exploration as striking a balance between approach and avoidance (Ahmadlou et al., 2021; Botta et al., 2020; Ogasawara et al., 2022). Our findings highlight the shared computational principles between human exploration in symbolic space and animal exploration of the physical environment and suggest that mechanisms involved in avoidance responses may also play a part in knowledge acquisition.

Finally, planning (Hunt et al., 2021), learning (Gureckis and Markant, 2012), and sensing (Gordon et al., 2014; Yang et al., 2016) are increasingly studied as active processes, situated within our environment and interacting with it. Understanding the complicated dynamics between agent and environment has been greatly facilitated by comparing behavior against the computational ideal of maximizing the amount of information observed (Oaksford and Chater, 1994; shman, 2019; Schwartenbeck et al., 2019). The findings we present here suggest a modification to this computational premise. Rather than trying to uncover as much information as possible, the goal of human exploration may be to maximize the amount of information retained in memory, by modulating the rate and order of observed information.

Methods

Data Collection and Participants

A sample of 298 participants was recruited via Amazon MTurk to participate in four sessions of the exploration task. They were paid $3.60 for each session and earned a bonus contingent on their test phase performance, adding up to $4.50 for the first session and $6 for later sessions. Additionally, a $2 bonus was paid out for completion of the fourth session. Participants were asked to complete the four sessions over the course of a week and were invited by email to each session after the first, as long as the data from their last session was not excluded according to the criteria we had specified (see below). All participants provided informed consent; all protocols were approved by the Columbia University Institutional Review Board.

The first session was terminated early for 89 participants due to recorded interactions with other applications during the experiment or failure to comply with instructions. An additional 32 sessions played by participants who had successfully completed the first session were excluded for the same reasons. One participant was excluded after reporting technical problems with stimulus presentation in the second session. Twenty-seven further sessions were excluded for failure to sample cards from both decks, a prerequisite for learning on which participants were instructed as part of the training. Altogether, data from 194 participants was included in the analyzed sample (120 female, 72 male, 2 other gender, average age 29.63, range 20-48). This sample included 194 first sessions, 156 second sessions, 129 third sessions, and 116 fourth sessions.

Before running this experiment, we pre-registered (Abir et al., 2021) a sample size of 190 participants satisfying our exclusion criteria. We chose this number to be three times larger than a preliminary sample of N=62 participants, which provided the dataset we used to develop our analysis approach and pipeline, and first identify exploration strategies as described above. Results for the preliminary sample are provided in figure supplements.

Task Design and Procedure

On each round of the exploration task participants were presented with a simple environment of four tables with two decks of cards on each table. Tables were distinguished by unique colorful patterns and decks by geometric symbols that did not repeat within an experimental session. The hidden side of each card was painted in one of two colors, with a unique color pair for each round. The proportion of colors in each deck were determined pseudo-randomly (see Supplementary Information), resulting in variability in the difference in proportion between each deck pair - the learning desideratum of this task.

At the beginning of each round, participants were first presented with the color pair for the round, and then with the table-deck assignments. Participants then had to pass a multiple-choice test on the table-deck assignment, making sure they remembered the structure of the task before proceeding to explore. Failing to get a perfect score on this test resulted in repeating this phase. The exploration phase then commenced. Trial structure for the exploration phase is depicted in **Figure 1b**. The lengths of the exploration phases varied from round to round. They were sampled from a geometric distribution with rate, shifted by 10 trials. The same list of round lengths was used for all participants, but their order was randomized.

Following the exploration phase, participants were tested on their learning. They were presented with the rewarding color for this round, and then had to indicate which deck had a greater proportion of that color on each table (**Figure 1c**). After answering this question for each of the four tables, they rated their confidence in each of the four choices on a 1-5 Likert scale. Participants were then told whether each of the test choices were correct, and the true color proportions for the two decks on each table were presented to them as 10 open cards.

The first session started with extensive instructions explaining the structure of each of the two phases of the task and clearly stating the learning goal. Participants were also instructed on the independence of color proportion within each deck pair, necessitating sampling from both decks to succeed in the task. The instructions also included training on how to make the relevant choices in each of the two stages. A quiz followed the instruction phase, and participants had to repeat reading the instructions if they had given the wrong response to any question on this quiz.

Each session started with a short practice round (12-19 trials). Data from this round was excluded from analysis. In the first session participants then played three more rounds and in later sessions five more rounds, for a total of 18 experimental rounds.

Data Analysis

Analysis was performed using Julia 1.4.2. Hierarchical regression models were fitted using the Stan probabilistic programming language 2.30.1 (Carpenter et al., 2017), using the interface supplied by the brims package version 2.16.1 (Burkner, 2017), running on top of R 4.1.2. The complete computing environment was packaged as a Docker image, which can be used to reproduce the entire analysis pipeline. Sequential sampling models were fitted on a separate Docker image (Chuan- Peng et al., 2022) containing HDDM 0.8 (Wiecki et al., 2013) running on top of python 3.8.8.

Bayesian Observer

Each of the three hypothesized strategies for exploration postulates a different summary statistic of prior learning as the driver of exploratory choice. To derive these summary statistics, we first had to construct a model of prior learning. We chose a simple Bayesian observer model (Behrens et al., 2007; Yang et al., 2016). Like our participants, this model's goal was to learn $\theta = sgn(\pi_1 - \pi_2)$ from observed outcomes $x_{0:t}$. It did so by placing a probabilistic prior over the value of each updating it after every observation according to Bayes' rule, and solving for θ using the rules of probability. The result is a posterior distribution capturing the agent's expectation of the value of θ, and their uncertainty about the expectation. This process is depicted in **Figure 2** for two tables and their matching pairs of decks.

This computation can be put into formulaic form as follows. At the beginning of a round, the Bayesian observer places a flat Beta distribution prior on the proportion of colors in each of the eight decks:

$$\pi_i \sim Beta(1, 1)$$

After observing a card, this prior would be updated to form a posterior distribution. Since the posterior of a Beta prior and a Bernoulli observation likelihood is also a Beta distribution, the posterior has a simple analytic form: after completing t trials, observing c_i cards of one color and $t - c_i$ cards of the other color, the posterior would

$$\pi_i | x_{0:t} \sim Beta(1 + c_i, 1 + t - c_i)$$

We can then find the probability that $\theta = 1$, i.e. that $\pi_1 > \pi_2$, by calculative the probability that π_2 is smaller than a given $\pi_1 = z$, and integrating over z, the possible values of π_1:

$$P(\theta = 1 | x_{0:t}) = \int_0^1 f_{\pi_1 | x_{0:t}}(z) F_{\pi_2 | x_{0:t}}(z) dz$$

Where f is the Beta probability density function, F is the Beta cumulative density function, and $x_{0:t}$ are observations thus far. We computed the value of this integral numerically using the Julia package QuadGK.jl. Finally, θ can only take two values, and so

$$P(\theta = -1|x_{0:t}) = 1 - P(\theta = 1|x_{0:t})$$

Computing Hypothesized Decision Variables

The theory of decision making defines a decision variable as the quantity evaluated by the decision maker in order to choose between two choice options (Shadlen and Kiani, 2013). The difficulty of the decision should scale with the absolute value of the decision variable. Each of the three hypothesized strategies is defined by a specific summary statistic of prior learning that might serve as the decision variable for an exploratory choice. The three summary statistics are given in **Figure 1e**.

Both EIG and uncertainty are derived from the uncertainty of the posterior for θ as defined above. We quantified uncertainty as the entropy of the posterior belief (MacKay, 1992; Oaksford and Chater, 1994; Yang et al., 2016).

$$H(\theta|x_{0:t}) = -\sum_{\theta=-1,1} P(\theta|x_{0:t}) ln P(\theta|x_{0:t})$$

Entropy takes the unit of nats, ranging from 0 should the participant be absolutely sure about the value of θ for both table choice options, to 0.69 when they know nothing about a table. This is the equivalentof1bitofinformation, were we to replace the natural logarithm with a base 2logarithm.

Estimating Multilevel Bayesian Models for Inference

The regression coefficients and PIs reported here were all estimated using multilevel regression models accounting for individual differences in behavior. We used regularizing priors for all coefficients of interest to facilitate robust estimation (Table 1). For RT data we selected informative priors for the intercepttterm in the regression (capturing the grand average of RTs)following established recommendations (Schad et al., 2021). For predicting choices, we used logistic regression, for confidence ratings we used ordinal-logistic regression, and for average RTs we used log normal regression. We estimated these models with Hamiltonian Monte Carlo implemented in the Stan probabilistic programming language using the R package brms. Three Monte-Carlo chains were run for each model, collecting 1000 samples each after a warm up period of at least 1000 samples (warm up was extended if convergence had not been reached). Sequential sampling models were estimated using slice sampling, implemented in the python package HDDM. Four Monte-Carlo chains were run for each model, collecting 2000 samples each after a warm up period of at least 2000 samples. Convergence for both model types was assessed using the $\hat{R}$ metric, and visual inspection of trace plots. R syntax formulae and coefficients for covariates for all models mentioned in the main text are reported in Supplementary Information.

Sequential Sampling Model of Reaction Times

To draw inference from participants' RTs we turned to the sequential sampling theory of deliberation and choice. This theory encompasses a family of models in which decisions arise through a process of sequential sampling that stops when the accumulation of evidence satisfies a threshold or bound (Palmer et al., 2005; Shadlen and Kiani, 2013). From this family of models we chose to use the drift diffusion model (DDM) to fit our data, as it is very well described and extensively studied (Ratcliff and McKoon, 2008; Shadlen and Kiani, 2013). The DDM explains RTs as the culmination of three interpretable terms. The first is the efficacy of a participant's thought process in furnishing relevant evidence for the decision - in our case the efficacy of calculating Δ-uncertainty (the drift rate in DDM parlance). The second term governs the participant's speed-accuracy tradeoff by determining how much evidence they require to commit to a decision. This can also be thought of as how long a participant is willing to deliberate when a decision is difficult(bound height). Finally, the portion of the RT not linked to the deliberation process is captured by a third term (non-decision time). Since behavior was considerably different when overall uncertainty was high, DDM models were fit excluding trials with total uncertainty above the participant's estimated threshold.

Model Evaluation

We compared the models of choice and RTs to alternative models, either reduced or expanded (see Supplementary Information). We used the LOO R package to perform approximate leave- one-out cross validation for models implemented in Stan. This method uses pareto-smoothed importance sampling to approximate cross validation in an efficient manner (Vehtari et al., 2017). Models implemented in HDDM were compared using the DIC metric. We also performed recovery analysis and posterior predictive checks for our models, making sure they capture the theoretically important qualitative features of the data.

Acknowledgements

We thank the Shohamy lab, Christopher A. Baldassano, and Gabriel M. Stine for their insightful discussion of the project. We are thankful for the support of the Stan user community, especially Matti Vuorre and Paul Burkner. We are grateful for funding support from the NSF (award #1822619 to D.S.), NIMH/NIH (#MH121093to D.S.)and the Templeton Foundation (#60844 to D.S.).

Additional Information

Author Contributions

Y.A., M.N.S, and D.S. designed research; Y.A. collected and analyzed data; M.N.S and D.S supervised.

Appendix 1

Full Details of Task and Procedure

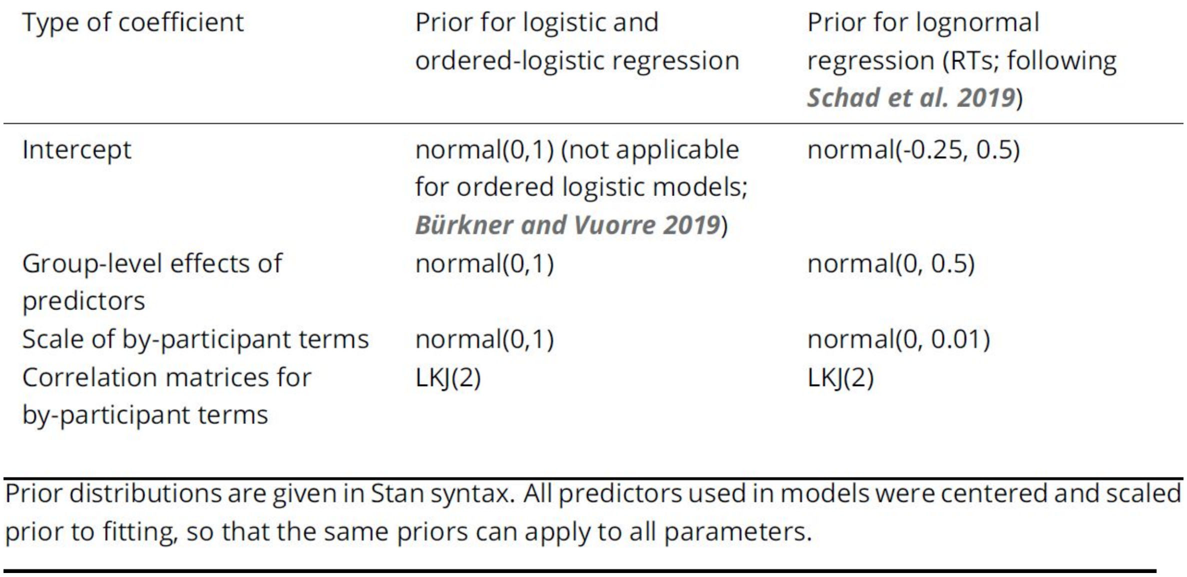
Recruitment

Participants were recruited from the pool of Amazon Mechanical Turk (MTurk) vetted by cloudresearch.com (Hauser et al., 2022; Litman et al., 2017). We further restricted enrollment to participants with an approval rating higher than 95%, and at least 100 prior jobs completed. Enrollment was also restricted to participants registered as being 18-35 years old. Participants were presented with an ad for a multi-session study, describing base pay and the performance-dependent bonus. The ad stated that we were only looking for participants who were willing to complete all four sessions, and that the task required undivided attention.

After accepting the task, participants were directed to a website running the experiment, which was coded using jsPsych 6.0.4 (De Leeuw, 2015).

Instructions and Training

At the beginning of the first session, participants were thoroughly instructed about the task, and practiced each of its phases. First, the two-phase structure and the learning goal were introduced. Participants then were shown an example table with two decks on it. They practiced choosing a deck on a single table. Participants were instructed that only the symbol on the deck determined its identity, while its location (which was randomly determined) did not matter. Participants then



Type of coefficient	Prior for logistic and ordered-logistic regression	Prior for lognormal regression (RTs; following ***Schad et al. 2019***)
Intercept	normal(0,1) (not applicable for ordered logistic models; ***Bürkner and Vuorre 2019***)	normal(-0.25, 0.5)
Group-level effects of predictors	normal(0,1)	normal(0, 0.5)
Scale of by-participant terms	normal(0,1)	normal(0, 0.01)
Correlation matrices for by-participant terms	LKJ(2)	LKJ(2)

Prior distributions are given in Stan syntax. All predictors used in models were centered and scaled prior to fitting, so that the same priors can apply to all parameters.

Table 1.

Regularizing priors used in regression models.

completed ten practice trials limited to choosing decks on a single table, with the goal of figuring out the difference in proportions of colors between the decks. In this practice, one deck had 7/10 cards of the rewarding color, and the other 9/10 cards. After practicing choosing decks, participants were told which had more of the rewarding color, and it was explained that the differences could be more difficult to figure out in the game itself. Next, participants were told that while both decks may have a majority of cards of the same color. As such, they would have to sample from both decks to learn which had more of each color. This point was demonstrated by presenting ten cards from two decks -and asking (i) which deck had a majority of color 1 cards (both did), (ii) which had a majority of color 2 cards (none did), (iii) which deck had more of color 1 than the other (one of the decks did), and (iv) which deck had more of color 2 than the other (the other deck did). A failure to give a correct answer on all four questions prompted a repetition of this section.

Participants were then introduced to choosing a table before making a deck choice. As practice, they were tasked with revealing a single card from each deck on two tables over 4 trials. Failing to do so resulted in participants having to repeat this section.

Next, the test phase was introduced. Participants were instructed that a particular deck had more of the rewarding color in it. They then had to choose that deck on the following test screen. Having successfully done so, they received the same visual feedback they would receive in the actual task - ten cards of each deck were presented to them, demonstrating the true proportion of colors in each deck. A message was displayed alongside this demonstration, stating the accuracy of their choice.

Before starting work on the main task, participants were reminded that the test phase could commence on any given trial. They then answered six multiple-choice questions about the structure of the task and their goal. If they got any of these wrong, they were instructed on the correct answer, and then had to repeat the quiz. After successfully completing the quiz, they played the whole task over a 20-trials-long practice round. For this practice, only two tables were included in the set of stimuli. Finally, after completing this practice round, they continued to play four more rounds of the full task with four tables.

Each of the following three sessions began with a short reminder of the instructions. Participants then completed five rounds of the task. The first round was always a short one (12-19 trials, these are the four lowest numbers drawn from the geometric distribution of round lengths), and was treated as a practice round in analysis, i.e. data from this round was discarded.

Procedure for Single Round

Familiarization

At the beginning of each round, participants were presented with each of the tables included in the round, and the two decks associated with each table. They were then tested on these associations: they were shown a table, and two decks next to it, and had to indicate which of the two belonged to the table. Failure to answer correctly on all eight trials resulted in a repetition of the introduction. Following the introductions of tables and decks, participants were shown the two card colors used in this round. They were reminded about their goal, and then commenced to play the learning phase.

Learning Phase

Each trial of the learning phase began with a 500ms period during which a fixation cross was displayed at the centre of the screen. Two tables were then presented as choice options. Choice options were chosen for each trial at random, and the left-right presentation of the two choice options was also determined randomly. Participants indicated their choice of table by pressing either the ‘d’ or ‘k’ keys. Next, the unchosen table was removed from the screen, and a frame with the same pattern as the chosen table was presented for a period of 1000 ms. This was followed by a 500ms presentation of a fixation cross at the center of this frame. Then, the two decks associated with the table appeared in the frame (deck location was determined randomly). Using the same keys, participants chose to reveal a card on one of the decks. A short animation was played showing the deck being shuffled, and a single card was flipped to reveal its color. The colored card remained on screen for 2400 ms. A 1700 ms inter-trial-interval followed each trial.

Test Phase

At the end of the learning phase, participants were instructed that they will now be tested on their learning, and were shown the color designated as the rewarding color for this round. They then chose the deck they believed had more of the rewarding color on each table. After making all four choices, participants were shown their choice on each table and asked to rate their confidence that their choice was indeed the correct one.

Memory Test

Following the confidence ratings, participants were tested on their memory of table and deck associations. They were shown each deck participating in the round, and had to indicate which of the four tables it belonged to. Participants received no feedback for this memory test.

Feedback

Next, participants received feedback for their test-phase choices. For each table they were shown ten cards drawn from each deck. The cards represented the true proportion of colors in the deck. They were reminded which deck they had chosen, and were told whether that was the correct choice. After observing this for each of the four tables, participants were told how much bonus money they earned in this round.

Debriefing

At the end of the experiment, after collecting demographic information, participants were asked about any strategy they may have implemented to remember the card colors better, to choose between decks and between tables in the learning phase, and between decks in the test phase. Lastly they were asked if anything in the instructions remained unclear. These responses were evaluated for any use of external aids, such as pen and paper, and for any technical difficulties.

Compliance and Attention Checks

We implemented several measures to incentivize participants to devote their undivided attention to the task. First, the task ran in full screen mode, and if participants chose to exit it, a warning message was shown explaining that this study only runs in fullscreen mode. Additionally, refreshing the webpage and right-clicking on it were disabled. Instances where the participant interacted with another application on their computer were recorded by jsPsych, and when this occurred a warning message was displayed on screen.

Participants had to make each learning-phase choice within 3000 ms. Failing to do so resulted in the display of a warning message asking them to choose more quickly. Additionally, if a reaction time (RT) of less than 250 ms was recorded on three consecutive choices, a warning message asking participants to comply with instructions was displayed on screen.

When more than ten warning messages of any kind had been displayed, the session terminated, and participants were asked to return the job to MTurk. Participants were paid for terminated sessions, but were not invited to following sessions.

Appendix 2

Preliminary Sample

Data Collection and Participants

A preliminary sample of 70 participants was recruited via MTurk to participate in four sessions of the task. Task design was identical to that later used for the pre-registered sample. Recruitment was not limited to the cloudresearch.com approved sample, since data collection predated the proliferation of fake worker accounts on MTurk (Hauser et al.,2022).

The first session was terminated early for 7 participants due to recorded interactions with other applications during the experiment, or failure to comply with instructions. An additional 8 sessions played by participants who had successfully completed the first session were excluded for the same reasons. Five further sessions were excluded for failure to sample cards from both decks, a prerequisite for learning on which participants were instructed as part of the training. Altogether, data from 62 participants was included in the analyzed sample (33 female, 28 male, 1 other gender, average age 29.10, range 20-38). This sample included 62 first sessions, 45 second sessions, 33 third sessions, and 29 fourth sessions.

Results

Results from the pre-registered sample largely replicate the results we first observed in the preliminary sample (see matching figure supplement for each figure). Two points of divergence are observed. In the preliminary sample, we didn't see a significant correlation between the tendency to avoid uncertainty when overall uncertainty is high and testperformance, while in the larger pre-registered sample we observed a positive significant correlation. This difference could be due to sample size, as the correlation is weak, we shouldn't over interpret it.

A second point of divergence regards the plotting of predicted RTs by test performance tertile. While both in the preliminary sample and the pre-registered sample we find that the bound height estimated by the DDM is predictive of test performance(Supplementary Table 12), in the preliminary sample this relationship does not translate to a monotonic rising of bound height when comparing participants by test performance tertile (Supplementary **Figure 4b**). We hesitate to interpret any non-linear relationship between the bound height and test performance, given the small size of the sample, the divergence from the pre-registered sample, and the complexity of the models involved.

Appendix 3

	Pre-registered sample (12379 trials, 194 participants)		Preliminary sample (3482 trials, 62 participants)		
Term	Median	95% PI	Median	95% PI	Units
		Predictors			
Intercept	0.86	[0.61, 1.11]	0.90	[0.47, 1.32]	logit
Final uncertainty	−5.59	[-6.25, -4.95]	−5.69	[-6.86, -4.68]	logit / nats
		Participant-wise variability			
SD of intercept	1.58	[1.38, 1.82]	1.45	[1.13, 1.89]	
SD of final uncertainty	7.49	[6.54, 8.61]	6.83	[5.33, 8.88]	
Correlation of intercept and final uncertainty	−0.68	[-0.84, -0.46]	−0.81	[-0.97, -0.40]	

The model can be summarized with the following R syntax formula: $accuracy \sim 0.5 + 0.5 * inv\_logit(1 + final\ uncertainty + (1 + final\ uncertainty|participant))$, where inv_logit is the inverse logit function. This functional form limits predicted accuracy between 0.5 and 1.0, since guessing-level accuracy on a two-alternative forced-choice test is 0.5. Since accuracy is a binary variable, this regression was fit with a Bernoulli likelihood.

Table 1.

Test Accuracy as a Function of the Final Uncertainty in the Exploration-Phase.

	Pre-registered sample (12007 trials, 194 participants)		Preliminary sample (3362 trials, 62 participants)		
Term	Median	95% PI	Median	95% PI	Units
Predictors					
Threshold 1	−2.22	[-2.44, -2.00]	−2.11	[-2.49, -1.72]	
Threshold 2	−0.44	[-0.67, -0.23]	−0.20	[-0.58, 0.17]	
Threshold 3	1.31	[1.08, 1.53]	1.60	[1.22, 1.98]	
Threshold 4	3.11	[2.88, 3.34]	3.58	[3.18, 3.98]	
Final uncertainty	−2.48	[-2.93, -2.06]	−2.32	[-3.11, -1.56]	logit / nats
Choice accuracy	1.09	[0.92, 1.27]	1.05	[0.71, 1.39]	logit
Final uncertainty × choice accuracy	−3.10	[-3.76, -2.46]	−3.20	[-4.60, -1.93]	logit / nats
Participant-wise variability					
SD of intercept	1.45	[1.31, 1.63]	1.43	[1.19, 1.73]	
SD of final uncertainty	2.10	[1.73, 2.52]	2.12	[1.39, 2.97]	
SD of choice accuracy	0.81	[0.64, 0.99]	0.84	[0.52, 1.25]	
SD of uncertainty × accuracy	2.11	[1.46, 2.82]	3.02	[1.58, 4.59]	
Correlation of intercept and uncertainty	0.23	[0.03, 0.41]	0.07	[-0.26, 0.39]	
Correlation of intercept and accuracy	−0.20	[-0.39, 0.01]	0.02	[-0.33, 0.40]	
Correlation of uncertainty and accuracy	−0.57	[-0.81, -0.30]	−0.51	[-0.83, -0.03]	
Correlation of intercept and uncertainty × accuracy	0.27	[-0.01, 0.52]	0.28	[-0.11, 0.62]	
Correlation of uncertainty and uncertainty × accuracy	0.68	[0.34, 0.91]	0.26	[-0.24, 0.70]	
Correlation of accuracy and uncertainty × accuracy	−0.84	[-0.95, -0.61]	−0.23	[-0.66, 0.35]	

The model can be summarized with the following R syntax formula: $confidence \sim 0 + final\,uncertainty * accuracy + (1 + final\,uncertainty * accuracy|PID)$. This model was fit as an ordered-logistic regression, with four threshold variables since confidence was rated on a 5-point likert scale (**?**).

Table 2.

Test Confidence as a Function of the Final Uncertainty in the Exploration-Phase.

	Pre-registered sample (146766 trials, 194 participants)		Preliminary sample (41009 trials, 62 participants)		
Term	Median	95% PI	Median	95% PI	Units
		Predictors			
Intercept	-0.04	[-0.10, 0.02]	-0.03	[-0.11, 0.06]	logit
Δ-uncertainty	0.89	[0.76, 1.03]	1.01	[0.70, 1.30]	logit / nat
		Participant-wise variability			
SD of intercept	0.39	[0.35, 0.43]	0.36	[0.30, 0.45]	
SD of Δ-uncertainty	0.89	[0.79, 1.00]	1.11	[0.91, 1.38]	
Correlation of intercept and Δ-uncertainty	-0.05	[-0.20, 0.12]	0.07	[-0.21, 0.31]	

The model can be summarized with the following R syntax formula: $table\ on\ right\ chosen \sim 1 + \Delta\ uncertainty + (1 + \Delta\ uncertainty|participant)$. This model was fit as a logistic regression.

Table 3.

Exploration-Phase Choices as a Function of Δ-Uncertainty

	Pre-registered sample (146766 trials, 194 participants)		Preliminary sample (41009 trials, 62 participants)		
Term	Median	95% PI	Median	95% PI	Units
		Predictors			
Intercept	-0.04	[-0.09, 0.02]	-0.03	[-0.12, 0.07]	logit
Δ-EIG	10.12	[8.00, 12.47]	12.50	[7.97, 16.91]	logit / nat
		Participant-wise variability			
SD of intercept	0.39	[0.35, 0.43]	0.35	[0.29, 0.43]	
SD of Δ-EIG	1.21	[1.07, 1.37]	1.48	[1.21, 1.82]	
Correlation of intercept and Δ-EIG	-0.02	[-0.17, 0.14]	-0.11	[-0.35, 0.16]	

The model can be summarized with the following R syntax formula: $tableonrightchosen \sim 1 + \Delta EIG + (1 + \Delta\ EIG|participant)$. This model was fit as a logistic regression.

Table 4.

Exploration-Phase Choices as a Function of Δ-EIG

	Pre-registered sample (146766 trials, 194 participants)		Preliminary sample (41009 trials, 62 participants)		
Term	Median	95% PI	Median	95% PI	Units
		Predictors			
Intercept	-0.04	[-0.10, 0.02]	-0.03	[-0.13, 0.07]	logit
Δ-exposure	-0.03	[-0.04, -0.02]	-0.03	[-0.05, -0.02]	logit / trial
		Participant-wise variability			
SD of intercept	0.39	[0.35, 0.43]	0.36	[0.29, 0.44]	
SD of Δ-exposure	0.05	[0.04, 0.06]	0.05	[0.04, 0.06]	
Correlation of intercept and Δ-exposure	0.16	[-0.01, 0.32]	-0.08	[-0.36, 0.21]	

The model can be summarized with the following R syntax formula: $table\ on\ right\ chosen \sim 1 + \Delta\ exposure + (1 + \Delta\ exposure|participant)$. This model was fit as a logistic regression.

Table 5.

Exploration-Phase Choices as a Function of Δ-Exposure

	Pre-registered sample (146766 trials, 194 participants)		Preliminary sample (41009 trials, 62 participants)		
Term	Median	95% PI	Median	95% PI	Units
		Predictors			
Intercept	-0.04	[-0.10, 0.02]	-0.03	[-0.13, 0.07]	logit
Δ-uncertainty	0.97	[0.83, 1.11]	1.12	[0.83, 1.42]	logit / nat
Δ-uncertainty × overall uncertainty	-428.44	[-536.60, -339.27]	-444.27	[-559.73, -353.23]	logit / nat^2
Transformed threshold α	2.52	[2.40, 2.64]	2.33	[2.17, 2.49]	a.u.
		Participant-wise variability			
SD of intercept	0.39	[0.35, 0.43]	0.36	[0.30, 0.44]	
SD of Δ-uncertainty	0.92	[0.81, 1.04]	1.09	[0.89, 1.35]	
SD of Δ-uncertainty × overall uncertainty	12.85	[0.56, 41.41]	8.97	[0.46, 28.35]	
SD of transformed threshold	0.45	[0.38, 0.54]	0.35	[0.26, 0.47]	

This model can be summarized with the following formula: $logit(P(tableonrightchosen)) = Intercept + b_1 * \Delta\ uncertainty + b_2 * (overall\ uncertainty - \omega) * step(overall\ uncertainty - \omega) * \Delta\ uncertainty$, where step is the step function, $\omega = -2ln(0.5) * inv\_logit(\alpha)$. The intercept and parameters b_1, b_2, and α all vary by participant. This model was fit as a logistic regression.

Table 6.

Exploration-Phase Choices as a Function of ΔΔ-uncertainty and Overall Uncertainty

Term	Pre-registered sample (194 participants)		Preliminary sample (62 participants)		Units
	Median	95% PI	Median	95% PI	
Predictors					
Intercept	1.56	[1.51, 1.61]	1.62	[1.53, 1.72]	logit
Approach tendency	2.96	[2.67, 3.25]	3.09	[2.65, 3.57]	logit^2 / nat
Participant-wise variability					

The model can be summarized with the following R syntax formula: $test\ accuracy \sim 1 + approach\ tendency$. For the tendency to approach uncertainty we computed the mean posterior approach parameter for each participant in the model described in ***Table 6***. The model described here was fit as a logistic regression with binomial likelihood.

Table 7.

Test Performance as a Function of the Tendency to Approach Uncertainty in Exploration

Term	Pre-registered sample (194 participants)		Preliminary sample (62 participants)		Units
	Median	95% PI	Median	95% PI	
Predictors					
Intercept	1.52	[1.48, 1.57]	1.59	[1.50, 1.68]	logit
Avoid tendency	1.18	[0.80, 1.58]	-0.52	[-1.20, 0.21]	nat^2
Participant-wise variability					

The model can be summarized with the following R syntax formula: $test\,accuracy \sim 1 + avoid\,tendency$. For the tendency to avoid uncertainty when overall uncertainty is high a we computed for each participant the area under the curve of the uncertainty approach / avoid graph, averaging across the posterior of the model described in ***Table 6***. The model described here was fit as a logistic regression with binomial likelihood.

Table 8.

Test Performance as a Function of Tendency to Approach Uncertainty in Exploration when Overall Uncertainty is High

	Pre-registered sample (113746 trials, 194 participants)		Preliminary sample (31205 trials, 62 participants)		
Term	Median	95% PI	Median	95% PI	Units
		Predictors			
B - bound height	0.74	[0.72, 0.76]	0.74	[0.71, 0.76]	
μ_0 - drift rate offset	-0.01	[-0.06, 0.03]	-0.01	[-0.08, 0.06]	
κ - dependence of drift rate on uncertainty	0.69	[0.58, 0.78]	0.78	[0.58, 0.99]	
t_{ND} - non-decision time	0.28	[0.26, 0.31]	0.27	[0.25, 0.31]	
		Participant-wise variability			
SD of B	0.12	[0.11, 0.14]	0.10	[0.08, 0.12]	
SD of μ_0	0.30	[0.27, 0.33]	0.25	[0.21, 0.31]	
SD of κ	0.66	[0.58, 0.74]	0.76	[0.62, 0.94]	
SD of t_{ND}	0.16	[0.14, 0.19]	0.12	[0.10, 0.15]	

We used a drift-diffusion model to formalize the dependence of RTs and choice on evidence. A drift-diffusion model is one variant in the sequential sampling family of models. The model posits that samples of momentary evidence are integrated over time. The expectation of the momentary evidence distribution is termed the drift rate μ, and its standard deviation is termed the diffusion coefficient. The decision is made when integrated evidence reaches an upper or lower bound ($\pm B$), whose sign determines the choice. Processes external to decision making are modelled by $t_N D$, a constant added to the RT. In this model μ is allowed to depend linearly on Δ-uncertainty, $\mu = \mu_0 + \kappa \cdot \Delta - uncertainty$, such that κ captures the dependence of drift rate on Δ-uncertainty, and μ_0 is a general bias to make rightward or leftward choices.

Prior to fitting the model to the data, we excluded trials for which overall uncertainty was above the threshold estimated for each participant by the model described in ***Table 6***. As we find qualitatively different choice behavior above the threshold, we couldn't justify modelling these trials together with the majority of trials. Fitting a piecewise regression DDM model was beyond the capabilities of current software.

Table 9.

Drift Diffusion Model of Exploration-Phase Choice and RTs

	Pre-registered sample		Preliminary sample		
	(113746 trials, 194 participants)		(31205 trials, 62 participants)		
Term	Median	95% PI	Median	95% PI	Units
		Predictors			
Intercept	0.33	[-0.34, 1.01]	0.23	[-0.86, 1.28]	
Final uncertainty	-0.87	[-0.98, -0.76]	-0.93	[-1.13, -0.74]	
B - bound height	1.46	[0.58, 2.34]	1.49	[0.05, 2.90]	
κ - dependence of drift rate on uncertainty	0.81	[0.58, 1.07]	0.88	[0.57, 1.21]	
		Participant-wise variability			
SD of intercept	0.87	[0.74, 1.01]	0.68	[0.47, 0.96]	
SD of final uncertainty	0.49	[0.39, 0.60]	0.44	[0.24, 0.67]	
Correlation of intercept and final uncertainty	-0.81	[-0.92, -0.66]	-0.83	[-0.98, -0.44]	

This model can be summarized with the following R syntax formula: $test\ accuracy \sim 1 + final\ uncertainty + B + \kappa + (1 + final\ uncertainty|participant)$. This model was fit as a logistic regression. As B and κ are parameters estimated from the model described in ***Table 9***, we took into account our error in measuring them when using them as predictors in this model. Thus, the posterior distribution for each participant's B and κ parameters was summarized as a mean and standard deviation. These summary statistics were used to approximate the posterior as a normal distribution from which a latent variable was drawn during the estimation of this model. This method propagates the uncertainty in the values of B and κ into the estimates reported here. Prior to using this method, we inspected the posteriors from the model summarized in ***Table 9***, and made sure the normal distribution is an adequate approximation for these posteriors.

Table 10.

Test Performance as a Function of Drift Diffusion Model Parameters for Exploration Phase

	Pre-registered sample (146766 trials, 194 participants)		Preliminary sample (41009 trials, 62 participants)		
Term	Median	95% PI	Median	95% PI	Units
		Predictors			
Intercept	-0.04	[-0.10, 0.02]	-0.03	[-0.14, 0.07]	logit
Δ-uncertainty	1.01	[0.86, 1.14]	1.16	[0.87, 1.44]	logit / nat
Δ-uncertainty × overall uncertainty	-326.14	[-468.61, -245.18]	-349.54	[-691.47, -251.13]	logit / nat^2
Transformed threshold	2.55	[2.40, 2.72]	2.36	[2.16, 2.67]	a.u.
Repeat choice on right	0.50	[0.42, 0.59]	0.57	[0.43, 0.72]	logit difference
		Participant-wise variability			
SD of intercept	0.40	[0.36, 0.44]	0.38	[0.31, 0.46]	
SD of Δ-uncertainty	0.90	[0.80, 1.02]	1.05	[0.86, 1.31]	
SD of Δ-uncertainty × overall uncertainty	14.21	[0.98, 41.45]	9.23	[0.46, 29.28]	
SD of transformed threshold	0.44	[0.36, 0.54]	0.35	[0.22, 0.50]	
SD of repeat choice	0.57	[0.51, 0.64]	0.53	[0.44, 0.66]	
Correlation of intercept and Δ-uncertainty	-0.06	[-0.21, 0.10]	0.05	[-0.21, 0.32]	
Correlation of intercept and Δ-uncertainty × overall uncertainty	-0.08	[-0.76, 0.72]	-0.01	[-0.76, 0.75]	
Correlation of Δ-uncertainty and Δ-uncertainty × overall uncertainty	0.01	[-0.69, 0.69]	-0.10	[-0.78, 0.70]	
Correlation of intercept and repeat choice	-0.16	[-0.30, -0.00]	-0.06	[-0.31, 0.22]	
Correlation of Δ-uncertainty and repeat choice	0.32	[0.17, 0.46]	0.11	[-0.18, 0.38]	
Correlation of Δ-uncertainty × overall uncertainty and repeat choice	0.38	[-0.67, 0.87]	0.00	[-0.76, 0.77]	
Correlation of intercept and threshold	-0.03	[-0.26, 0.22]	0.31	[-0.09, 0.64]	
Correlation of Δ-uncertainty and threshold	-0.35	[-0.52, -0.16]	0.09	[-0.27, 0.43]	
Correlation of Δ-uncertainty × overall uncertainty and threshold	-0.42	[-0.88, 0.61]	-0.10	[-0.79, 0.70]	
Correlation of repeat choice and threshold	-0.60	[-0.74, -0.43]	-0.38	[-0.64, -0.05]	

Table 11.

Exploration-phase Choices as a Function of Δ-Uncertainty, Overall Uncertainty, and Side of Repeat Option

	Pre-registered sample (113746 trials, 194 participants)		Preliminary sample (31205 trials, 62 participants)		
Term	Median	95% PI	Median	95% PI	Units
		Predictors			
B_0 - average bound height	0.75	[0.73, 0.76]	0.74	[0.72, 0.77]	
B_{repeat} - difference in bound height between repeat and switch chosen	-0.05	[-0.05, -0.04]	-0.04	[-0.06, -0.02]	
μ_0 - drift rate offset	-0.01	[-0.06, 0.03]	-0.01	[-0.08, 0.05]	
κ - dependence of drift rate on uncertainty	0.70	[0.61, 0.80]	0.81	[0.60, 1.01]	
κ_{repeat} - difference in dependence between repeat and switch chosen	-0.32	[-0.43, -0.22]	-0.28	[-0.49, -0.08]	
t_{ND} - non-decision time	0.28	[0.26, 0.31]	0.28	[0.25, 0.31]	
		Participant-wise variability			
SD of B_0	0.12	[0.11, 0.14]	0.10	[0.08, 0.12]	
SD of B_{repeat}	0.05	[0.04, 0.05]	0.05	[0.04, 0.07]	
SD of μ_0	0.30	[0.27, 0.33]	0.26	[0.21, 0.31]	
SD of κ	0.64	[0.57, 0.72]	0.74	[0.60, 0.92]	
SD of κ_{repeat}	0.50	[0.40, 0.60]	0.58	[0.39, 0.80]	
SD of t_{ND}	0.16	[0.14, 0.19]	0.12	[0.10, 0.15]	

We refit the DDM described in ***Table 9***, allowing both B and κ to vary by whether the choice was a repeat or switch choice (see main text for definition).

Table 12.

Drift Diffusion Model of Exploration-Phase Choice and RTs, Differentiating between Repeat and Switch Choices

	Pre-registered sample (194 participants)		Preliminary sample (62 participants)		
Term	Median	95% PI	Median	95% PI	Units
		Predictors			
Intercept	1.52	[1.47, 1.56]	1.59	[1.50, 1.67]	logit
Tendency to repeat	0.09	[0.07, 0.11]	0.11	[0.06, 0.16]	logit / logit
		Participant-wise variability			

The model can be summarized with the following R syntax formula: $test\ accuracy \sim 1 + tendency\ to\ repeat$. For tendency to repeat we computed the mean posterior parameter for each participant in the model described in ***Table 11***. This model was fit as a logistic regression with binomial likelihood.

Table 13.

Test Performance as a Function of the Tendency to Repeat Exploration-Phase Choices

	Pre-registered sample (126848 trials, 194 participants)		Preliminary sample (35264 trials, 62 participants)		
Term	Median	95% PI	Median	95% PI	Units
		Predictors			
Intercept	−0.34	[-0.38, -0.30]	−0.35	[-0.42, -0.30]	log s
Memory lag	0.02	[0.02, 0.03]	0.03	[0.02, 0.03]	log s / trial
Repeat choice on right	−0.05	[-0.06, -0.04]	−0.05	[-0.07, -0.03]	log s difference
Memory lag × repeat	0.02	[0.02, 0.03]	0.02	[0.02, 0.03]	1 / trial
		Participant-wise variability			
SD of intercept	0.29	[0.26, 0.31]	0.23	[0.19, 0.27]	
SD of memory lag	0.02	[0.01, 0.02]	0.02	[0.01, 0.02]	
SD of repeat choice on right	0.06	[0.05, 0.07]	0.07	[0.06, 0.09]	
SD of memory lag × repeat	0.02	[0.01, 0.02]	0.02	[0.01, 0.03]	
Correlation of intercept and memory lag	0.41	[0.24, 0.56]	0.19	[-0.11, 0.46]	
Correlation of intercept and repeat	−0.27	[-0.42, -0.11]	−0.35	[-0.57, -0.07]	
Correlation of memory lag and repeat	−0.70	[-0.83, -0.53]	−0.27	[-0.57, 0.07]	
Correlation of intercept and memory lag × repeat	0.27	[0.04, 0.48]	0.26	[-0.17, 0.65]	
Correlation of memory lag and memory lag × repeat	0.75	[0.51, 0.90]	0.21	[-0.28, 0.65]	
Correlation of repeat and memory lag × repeat	−0.91	[-0.98, -0.77]	−0.74	[-0.94, -0.36]	

The model can be summarized with the following R syntax formula: $log\ RT \sim 1 + memory\ lag * repeatonright + (1 + memorylag \cdot repeatonright | participant)$. This model was fit as a lognormal regression.

Table 14.

Exploration-Phase RTs as a Function of Memory Lag and Side of Repeat Option

	Pre-registered sample (126973 trials, 194 participants)		Preliminary sample (35304 trials, 62 participants)		
Term	Median	95% PI	Median	95% PI	Units
		Predictors			
Intercept	−0.03	[-0.09, 0.02]	−0.02	[-0.12, 0.08]	logit
Δ-uncertainty	1.03	[0.89, 1.17]	1.16	[0.87, 1.43]	logit / nat
Memory lag	−0.01	[-0.02, 0.00]	0.01	[-0.00, 0.02]	logit / trial
Repeat choice on right	0.45	[0.37, 0.52]	0.50	[0.37, 0.63]	logit difference
Δ-uncertainty × memory lag	−0.08	[-0.11, -0.04]	−0.14	[-0.20, -0.07]	logit / nat * trial
Memory lag × repeat	−0.13	[-0.15, -0.11]	−0.08	[-0.12, -0.04]	1 / trial
		Participant-wise variability			
SD of intercept	0.40	[0.36, 0.45]	0.37	[0.31, 0.45]	
SD of Δ-uncertainty	0.92	[0.81, 1.04]	1.06	[0.87, 1.32]	
SD of memory lag	0.03	[0.02, 0.04]	0.02	[0.00, 0.04]	
SD of repeat choice on right	0.49	[0.44, 0.55]	0.45	[0.36, 0.57]	
SD of Δ-uncertainty × memory lag	0.13	[0.08, 0.17]	0.07	[0.00, 0.19]	
SD of memory lag × repeat	0.10	[0.08, 0.13]	0.11	[0.07, 0.16]	

The model can be summarized with the following R syntax formula: $table\ on\ right\ chosen \sim 1 + \Delta\text{-}uncertainty \cdot memory\ lag + memory\ lag : repeat\ on\ right + (1 + \Delta\text{-}uncertainty \cdot memory\ lag + memory\ lag : repeat\ on\ right | participant)$. This model was fit as a logistic regression. For brevity, the correlations in participant-wise variability are emitted from this table.

Table 15.

Exploration-Phase Choices as a Function of ΔΔ-Uncertainty, Memory Lag, and Side of Repeat Option

	Pre-registered sample		Preliminary sample		
	(146766 trials, 194 participants)		(41009 trials, 62 participants)		
Term	Median	95% PI	Median	95% PI	Units
		Predictors			
Intercept	-0.04	[-0.09, 0.02]	-0.03	[-0.12, 0.06]	logit
Δ-uncertainty	1.00	[0.85, 1.14]	1.16	[0.85, 1.46]	logit / nat
Δ-uncertainty × overall uncertainty	-480.86	[-628.28, -379.29]	-450.20	[-568.82, -356.27]	logit / nat²
Transformed threshold	2.56	[2.43, 2.69]	2.32	[2.18, 2.48]	a.u.
Δ-uncertainty × trial #	0.00	[-0.00, 0.00]	0.00	[-0.01, 0.00]	logit / nat × trial
		Participant-wise variability			
SD of intercept	0.39	[0.35, 0.43]	0.36	[0.30, 0.45]	
SD of Δ-uncertainty	0.94	[0.83, 1.06]	1.13	[0.93, 1.39]	
SD of Δ-uncertainty × overall uncertainty	10.68	[0.52, 36.54]	8.87	[0.45, 28.51]	
SD of transformed threshold	0.43	[0.36, 0.51]	0.34	[0.25, 0.46]	
SD Δ-uncertainty × trial #	0.02	[0.01, 0.02]	0.02	[0.01, 0.02]	

We refit the piecewise regression model described in ***Table 6***, accounting for a possible interaction between Δ-uncertainty and trial number. We find no significant interaction in the pre-registered sample, nor the preliminary sample. All other terms in the model remained practically the same. The model can be summarized with the following formula: $logit(P(table on right chosen)) = Intercept + b1 \cdot \Delta\text{-}uncertainty + b2 \cdot (overall\ uncertainty - \omega) \cdot step(overall\ uncertainty - \omega) \cdot \Delta\text{-}uncertainty + b4 \cdot trial\ \# \cdot \Delta\text{-}uncertainty$, where step is the step function, $\omega = -2ln(0.5) \cdot inv\_logit(\alpha)$. The intercept and parameters b1, b2, b4, and α all vary by participant. This model was fit as a logistic regression.

Table 16.

Exploration-Phase Choices as a Function of Δ-Uncertainty, Overall Uncertainty, and Trial Number

Parameters varying by repeat / switch choice	Pre-registered sample DIC	Preliminary sample DIC
None	218,877.63	59,365.40
κ - the dependence of RT on Δ-uncertainty	218,666.99	59,311.25
B - Bound height	217,928.89	59,096.38
Both κ and B	217,669.43	59,046.67

The model reported in ***Table 12*** captures the tendency to repeat previous choices by allowing both the dependence of RT on Δ-uncertainty and the bound height parameters to vary by whether the choice was a repeat or switch choice (last row in this table). Here, it is compared against the simpler models nested within it. For both samples, the full model is favored over the partial models, as is indicated by lower deviance information criterion (DIC) values. DIC values are derived from the likelihood of the data given estimated parameters, and the effective number of parameters in the model. Absolute values of DIC depend on sample size and the attributes of the noise distribution. Accordingly, DIC values should only be compared between models fit to the same dataset.

Table 17.

Model Comparison for Sequential Sampling Models of the Tendency to Repeat Previous Choices

References

Abir Y, Shadlen MN, Shohamy D (2021) **Y Abir, MN Shadlen, D Shohamy, Memory-based incremental exploration in a stochastic environment; 2021. https://aspredicted.org/hx6gj.pdf.** *Memory-based incremental exploration in a stochastic environment*

Ahmadlou M *et al.* (2021) **cell type-specific cortico-subcortical brain circuit for investigatory and novelty-seeking behavior.** *Science*

Anderson JR. (1990) **The adaptive character of thought.** *Psychology Press;*

Auer P. (2002) **Using confidence bounds for exploitation-exploration trade-offs.** *Journal of Machine Learning Research* :397–422

Badia AP, Piot B, Kapturowski S, Sprechmann P, Vitvitskyi A, Guo ZD, Blundell C. (2020) **Agent57: Outperforming the Atari Human Benchmark** *Proceedings of the 37th International Conference on Machine Learning PMLR;* :507–517

Bavard S, Rustichini A, Palminteri S. (2021) **Two sides of the same coin: Beneficial and detrimental consequences of range adaptation in human reinforcement learning** *Science Advances* https://doi.org/10.1126/sciadv.abe0340,

Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) **Learning the value of information in an uncertain world** *Nature Neuroscience* :1214–1221 https://doi.org/10.1038/nn1954,

Bellemare M, Srinivasan S, Ostrovski G, Schaul T, Saxton D, Munos R. (2016) **Unifying count-based exploration and intrinsic motivation.** *Advances in neural information processing systems*

Botta P, Fushiki A, Vicente AM, Hammond LA, Mosberger AC, Gerfen CR, Peterka D, Costa RM. (2020) **An Amygdala Circuit Mediates Experience-Dependent Momentary Arrests during Exploration** *Cell* **183**:605–619 https://doi.org/10.1016/j.cell.2020.09.023.

Brown VM, Hallquist MN, Frank MJ, Dombrovski AY. (2022) **Humans adaptively resolve the explore-exploit dilemma under cognitive constraints: Evidence from a multi-armed bandit task** *Cognition* **229** https://doi.org/10.1016/j.cognition.2022.105233.

Bürkner PC (2017) **sbrims: An R Package for Bayesian Multilevel Models Using Stan** *Journal of Statistical Software* **80**:1–28 https://doi.org/10.18637/jss.v080.i01.

Bürkner PC, Vuorre M. (2019) **Ordinal regression models in psychology: A tutorial** *Advances in Methods and Practices in Psychological Science* **2**:77–101

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. (2017) **Stan: A Probabilistic Programming Language** *Journal of Statistical Software* https://doi.org/10.18637/jss.v076.i01,

Carrillo JD, Mariotti T. (2000) **Strategic ignorance as a self-disciplining device** *The Review of Economic Studies* :529–544

Chater N, Oaksford M. (1999) **Ten years of the rational analysis of cognition** *Trends in Cognitive Sciences* **3**:57–65 https://doi.org/10.1016/S1364-6613(98)01273-X

Chuan-Peng H, Geng H, Zhang L, Fengler A, Frank M, Zhang RY, Hitchhiker's A (2022) **Guide to Bayesian Hierarchical Drift-Diffusion Modeling with docker HDDM** *PsyArXiv;* https://doi.org/10.31234/osf.io/6uzga

Cohen JD, McClure SM, Yu AJ (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**:933–942 https://doi.org/10.1098/rstb.2007.2098.

Collins AGE, Frank MJ (2012) **How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis** *European Journal of Neuroscience* **35**:1024–1035 https://doi.org/10.1111/j.1460-9568.2011.07980.x

Collins AGE, Ciullo B, Frank MJ, Badre D. (2017) **Working memory load strengthens reward prediction errors** *Journal of Neuroscience* **37**:4332–4342 https://doi.org/10.1523/JNEUROSCI.2700-16.2017.

Daw ND, Niv Y, Dayan P. (2005) **Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control** *Nature neuroscience* **8**:1704–1711

Daw ND, O'doherty JP, Dayan P, Seymour B, Dolan RJ. (2006) **Cortical substrates for exploratory decisions in humans** *Nature* :876–879

De Leeuw JR. (2015) **jsPsych: A JavaScript library for creating behavioral experiments in a Web browser** *Behavior research methods* **47**:1–12

Duncan K, Semmler A, Shohamy D. (2019) **Modulating the Use of Multiple Memory Systems in Value-based Decisions with Contextual Novelty** *Journal of Cognitive Neuroscience* **31**:1455–1467 https://doi.org/10.1162/jocn_a_01447

Eilam D, Golani I. (1989) **Home base behavior of rats (Rattus norvegicus) exploring a novel environment** *Behavioural Brain Research* **34**:199–211 https://doi.org/10.1016/S0166-4328

Ellsberg D. (1961) **Risk, ambiguity, and the Savage axioms** *The quarterly journal of economics* :643–669

Fox CR, Tversky A. (1995) **Ambiguity a version and comparative ignorance** *The quarterly journal of economics* :585–603

Gigerenzer G, Garcia-Retamero R. (2017) **Cassandra's regret: The psychology of not wanting to know** *Psychological review*

Glickman SE, Sroges RW. (1966) **Curiosity in zoo animals** *Behaviour* :151–187

Golman R, Hagmann D, Loewenstein G. (2017) **Information avoidance** *Journal of economic literature* **55**:96–135

Gordon G, Fonio E, Ahissar E. (2014) **Emergent Exploration via Novelty Management** *Journal of Neuroscience* **34**:12646–12661 https://doi.org/10.1523/JNEUROSCI.1872-14.2014.

Gureckis TM, Markant DB. (2012) **Self-directed learning: A cognitive and computational perspective** *Perspectives on Psychological Science* **7**:464–481

Hartley CA (2022) **How do natural environments shape adaptive cognition across the lifespan?** *Trends in Cognitive Sciences* **26**:1029–1030 https://doi.org/10.1016/j.tics.2022.10.002.

Hauser DJ, Moss AJ, Rosenzweig C, Jaffe SN, Robinson J, Litman L. (2022) **Evaluating Cloud Research's Approved Group as a solution for problematic data quality on MTurk** *Behavior Research Methods* https://doi.org/10.3758/s13428-022-01999-x

Hunt LT *et al.* (2021) **Formalizing planning and information search in naturalistic decision-making** *Nature Neuroscience* **24**:1051–1064 https://doi.org/10.1038/s41593-021-00866-w,

Knowlton BJ, Mangels JA, Squire LR. (1996) **A Neostriatal Habit Learning System in Humans** *Science* **273**:1399–1402 https://doi.org/10.1126/science.273.5280.1399

Lieder F, Griffiths TL. (2020) **Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources** *Behavioral and Brain Sciences* https://doi.org/10.1017/S0140525X1900061X,

Litman L, Robinson J, Abberbock T. (2017) **TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences** *Behavior Research Methods* **49**:433–442 https://doi.org/10.3758/s13428-016-0727-z

MacKay DJC. (1992) **Information-based objective functions for active data selection** *Neural computation* **4**:590–604

Markant DB, Gureckis TM. (2014) **A preference for the unpredictable over the informative during self-directed learning** *Proceedings of the 36th Annual Conference of the Cognitive Science Society*

Nicholas J, Daw ND, Shohamy D. (2022) **Uncertainty alters the balance between incremental learning and episodic memory** *eLife* https://doi.org/10.7554/eLife.81679

Oaksford M, Chater N. (1994) **A Rational Analysis of the Selection Task as Optimal Data Selection** *Psychological Review* **101**:608–631 https://doi.org/10.1037/0033-295X.101.4.608.

Ogasawara T, Sogukpinar F, Zhang K, Feng YY, Pai J, Jezzini A, Monosov IE. (2022) **A primate temporal cortex-zona incerta pathway for novelty seeking** *Nature Neuroscience* :50–60 https://doi.org/10.1038/s41593-021-00950-1,

Palmer J, Huk AC, Shadlen MN. (2005) **The effect of stimulus strength on the speed and accuracy of a perceptual decision** *Journal of vision* **5**

Pathak D, Agrawal P, Efros AA, Darrell T. (2017) **Curiosity-driven Exploration by Self-supervised Prediction** *Proceedings of the 34th International Conference on Machine Learning PMLR* :2778–2787

Petitet P, Attaallah B, Manohar SG, Husain M. (2021) **The computational cost of active information sampling before decision-making under uncertainty** *Nature Human Behaviour* :935–946 https://doi.org/10.1038/s41562-021-01116-6,

Plonsky O, Teodorescu K, Erev I. (2015) **Reliance on small samples, the wavy recency effect, and similarity-based learning** *Psychological review* **122**

Poldrack RA, Clark J, Paré-Blagoev EJ, Shohamy D, Creso Moyano J, Myers C, Gluck MA. (2001) **Interactive memory systemsinthehumanbrain** *Nature* **4**:546–550 https://doi.org/10.1038/35107080.

Raposo D, Ritter S, Santoro A, Wayne G, Weber T, Botvinick M, van Hasselt H, Song F (2021) **Synthetic Returns for Long-Term Credit Assignment** *arXiv;*

Ratcliff R, McKoon G. (2008) **The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks** *Neural Computation* **20**:873–922 https://doi.org/10.1162/neco.2008.12-06-420.

Rothe A, Lake BM, Gureckis TM. (2018) **Do People Ask Good Questions?** *Computational Brain & Behavior* **1**:69–89 https://doi.org/10.1007/s42113-018-0005-5.

Ruggeri A, Sim ZL, Xu F. (2017) **Whyis Toma late to school again?" Preschoolers identify the most informative questions** *Developmental psychology*

Schad DJ, Betancourt M, Vasishth S. (2019) **Toward a principled Bayesian workflow in cognitive science** *arXiv preprint arXiv:190412765*

Schad DJ, Betancourt M, Vasishth S. (2021) **Toward a principled Bayesian workflow in cognitive science** *Psychological methods*

Schulz E, Gershman SJ. (2019) **The algorithmic architecture of exploration in the human brain** *Current Opinion in Neurobiology* **55**:7–14 https://doi.org/10.1016/j.conb.2018.11.003.

Schwartenbeck P, Passecker J, Hauser TU, FitzGerald TH, Kronbichler M, Friston KJ. (2019) **Computational mechanisms of curiosity and goal-directed exploration** *eLife* **8**:1–45 https://doi.org/10.7554/eLife.41703.

Sebastiani P, Wynn HP. (2000) **Maximum entropy sampling and optimal Bayesian experimental design** *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**:145–157

Shadlen MN, Kiani R. (2013) **Decision making as a window on cognition** *Neuron* **80**:791–806 https://doi.org/10.1016/j.neuron.2013.10.047

Shadlen MN, Shohamy D. (2016) **Decision Making and Sequential Sampling from Memory** *Neuron* **90**:927–939 https://doi.org/10.1016/j.neuron.2016.04.036.

Shafir E. (1994) **Uncertainty and the difficulty of thinking through disjunctions** *Cognition* **50**:403–430 https://doi.org/10.1016/0010-0277(94)90038-8

Shushruth S, Zylberberg A, Shadlen MN. (2022) **Sequential sampling from memory underlies action selection during abstract decision-making** *Current Biology* **32**:1949–1960

Song M, Bnaya Z, Ma WJ. (2019) **Sources of suboptimality in a minimalistic explore-exploit task** *Nature Human Behaviour* :361–368 https://doi.org/10.1038/s41562-018-0526-x,

Speekenbrink M, Konstantinidis E. (2015) **Uncertainty and Exploration in a Restless Bandit Problem** *Topics in Cognitive Science* **7**:351–367 https://doi.org/10.1111/tops.12145

Sutton RS, Barto AG. (2018) **Reinforcement learning: An introduction, 2nd ed. Reinforcement learning: An introduction, 2nd ed** :xxii–526

Trudel N, Scholl J, Klein-Flügge MC, Fouragnan E, Tankelevitch L, Wittmann MK, Rushworth MFS. (2020) **Polarity of uncertainty representation during exploration and exploitation in ventromedial prefrontal cortex** *Nature Human Behaviour* https://doi.org/10.1038/s41562-020-0929-3

Tversky A, Edwards W. (1966) **Information versus reward in binary choices** *Journal of Experimental Psychology* :680–683 https://doi.org/10.1037/h0023123,

Vehtari A, Gelman A, Gabry J. (2017) **Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC** *Statistics and Computing* **27**:1413–1432 https://doi.org/10.1007/s11222-016-9696-4

Waskom ML, Okazawa G, Kiani R. (2019) **Designing and Interpreting Psychophysical Investigations of Cognition** *Neuron* :100–112

Wiecki T, Sofer I, Frank M. (2013) **HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python** *Frontiers in Neuroinformatics* **7**

Wilson RC, Geana A, White JM, Ludvig EA, Cohen JD. (2014) **Humans use directed and random exploration to solve the explore-exploit dilemma** *Journal of Experimental Psychology: General*

Wu CM, Schulz E, Pleskac TJ, Speekenbrink M. (2022) **Time pressure changes how people explore and respond to uncertainty** *Scientific Reports* https://doi.org/10.1038/s41598-022-07901-1,

Wulff DU, Mergenthaler-Canseco M, Hertwig R. (2018) **A meta-analytic review of two modes of learning and the description-experience gap** *Psychological bulletin* **144**

Yang SCH, Lengyel M, Wolpert DM. (2016) **Active sensing in the categorization of visual patterns** *eLife* **5**:1–22 https://doi.org/10.7554/elife.12215.

Zylberberg A. (2021) **Decision prioritization and causal reasoning in decision hierarchies** *PLOS Computational Biology* https://doi.org/10.1371/journal.pcbi.1009688,

Article and author information

Yaniv Abir
Department of Psychology, Columbia University, New York, NY, USA
For correspondence: yaniv.abir@columbia.edu

Michael N. Shadlen
Zuckerman Mind Brain Behavior Institute, and Kavli Institute for Brain Science, Columbia University, New York, NY, USA, Department of Neuroscience and Howard Hughes Medical Institute, Columbia University, New York, NY, USA

Daphna Shohamy
Department of Psychology, Columbia University, New York, NY, USA, Zuckerman Mind Brain Behavior Institute, and Kavli Institute for Brain Science, Columbia University, New York, NY, USA
For correspondence: ds2619@columbia.edu

Copyright

Editors

Reviewing Editor
Claire Gillan
Trinity College Dublin, Dublin, Ireland

Senior Editor
Floris de Lange
Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands

Reviewer #1 (Public Review):

This manuscript reports on the behavior of participants playing a game to measure exploration. Specifically, participants completed a task with blocks of exploratory choices (choosing between two 'tables', and within each table, two 'card decks', each of which had a specific probability of showing cards with one color versus another) and test choices, where participants were asked to choose which of the two decks per table had a higher likelihood of one color. Blocks differed on how long (how many trials) the exploration phase lasted. Participants' choices were fit to increasingly complex models of next-trial exploration. Participants' choices were best fit by an intermediate model where the difference in uncertainty between tables influenced the choice. Next, the authors investigated factors affecting whether participants sought out or avoided uncertainty, their choice reaction times, and the relationship of these measures with performance during the test phase of each block. Participants were uncertainty-seeking (exploratory) under most levels of overall uncertainty but became less uncertainty-seeking at high levels of total uncertainty. Participants with a stronger tendency to approach uncertainty at lower levels of total uncertainty were more accurate in the test phase, while the tendency to avoid uncertainty when total uncertainty was high was also weakly positively related to test accuracy. In terms of reaction times, participants whose reaction times were more related to the level of uncertainty, and who deliberated longer, performed better. The individual tendency to repeat choices was related to avoidance of uncertainty under high total uncertainty and better test performance. Lastly, choices made after a longer lag were less affected by these measures.

The authors note that their paradigm, which does not provide immediate rewarding feedback, is novel. However, the resulting behavior appears similar to other exploratory learning tasks, so it's unclear what this task design adds - besides perhaps showing that exploratory behavior is similar across types of reward environments. Several papers have shown that cognitive constraints modulate exploration (PMIDs: 30667262, 24664860, 35917612, 35260717); although this paper provides novel insights, it does not situate its findings in the context of this prior literature. As a result, what it adds to the literature is difficult to discern.

Other methodological questions include whether the same model provides the best fit for all participants and whether possible individual differences in models used relate to individual differences in exploration and performance; how some analyses were carried out that currently lack sufficient detail in the manuscript; and how the two stages of choice behavior (tables versus card decks) were accounted for in the analyses.

https://doi.org/10.7554/eLife.94231.1.sa1

Reviewer #2 (Public Review):

Summary:
This paper focuses on an interesting question that has puzzled psychologists for decades, that is, why do people demonstrate a mix of uncertainty approach and avoidance behavior, given the fact that reducing uncertainty could always gain information and seems beneficial? This paper designed a novel task to demonstrate behavioral signatures of uncertainty approaching and avoidance during the exploration phase within the same task at both a within-subject and between-subject level. On the algorithmic level, this paper compared four different implementations of uncertainty-guided exploration and found that the model sensitive to relative uncertainty provides the best fit for human behavior compared to its counterparts using expected information gain or past exposure. This paper then links people's uncertainty attitude with accuracy and finds that uncertainty avoidance during exploration does not impair task performance, implying that uncertainty avoidance may be the output of a resource-rational decision-making process. To examine this account, this paper uses reaction time as an independent proxy of costly deliberation and shows that people deliberate shorter when engaging in repetitive choice, which presumably saves cognitive resources. Finally, the paper shows that people's tendency to engage in repetitive choice correlates with their tendency to avoid uncertainty, which supports the argument that avoiding uncertainty could be a strategy developed under the constraint of limited cognitive resources.

Strengths:
One of the highlights of this paper, as mentioned in the previous paragraph, is that the authors can establish the existence of the uncertainty approach and avoidance behavior within the same task whereas previous work usually focuses on one of them. This dissociation allows the authors to examine what situational factor is related to the emergence of the act of avoiding uncertainty, and extract parameters describing participants' attitude towards uncertainty during baseline as well as during situations where uncertainty avoidance is more common. Besides documenting the existence of uncertainty avoidance behavior, this paper also tried to explain this behavior by proposing under the resource rational framework and has carefully quantified different aspects (e.g., accuracy; choice speed) of participants' behavior as well as examined their relationships. Though more experiments are needed to fully understand human uncertainty avoidance behavior, this paper has provided both empirical and theoretical contributions toward a mechanistic understanding of how people balance approaching and avoiding uncertainty.

Weaknesses:
I have a couple of concerns related to this paper. First, there seems to exist an anti-correlation between total uncertainty and absolute relative uncertainty (Figure 5 panel C, \delta uncertainty is restricted to a small range when total uncertainty is high). It seems to be a natural product of the exploration process since the high total uncertainty phase is usually the period where the participant knows little about either option, leading to a less distinguishable relative uncertainty. However, it remains unknown whether the documented uncertainty avoidance still applies when extrapolating to larger absolute relative uncertainty. It would be great if the experiment allows for a manipulation of uncertainty in the middle of the experiment (e.g., introducing a new deck/informing that one deck has been updated). Relatedly, the current 'threshold' of uncertainty avoidance behavior, if I understand correctly, is found by empirically fitting participants' data. This brings the question: can we predict when people will demonstrate uncertainty avoidance behavior before collecting any data? Or, is it possible that by measuring some metrics related to cognitive cost sensitivity, we could predict the proportion of choices that participants will show uncertainty-avoidant behavior?

Finally, regarding the analysis of different behavior patterns in the game, it seems that the authors try to link repetitive behavior, uncertainty attitude, and accuracy together by testing the correlation between the two of them. I wonder whether other multivariate statistical methods e.g., mediation analysis, will be better suited for this purpose.

https://doi.org/10.7554/eLife.94231.1.sa0