Pupil Size Encodes Uncertainty during Exploration

Haoxue Fan[1], Taylor Burke[1], Deshawn Chatman Sambrano[1], Emily Dial[1], Elizabeth A. Phelps[1], and Samuel J. Gershman[1,2]

Abstract

■ Exploration is an important part of decision making and is crucial to maximizing long-term rewards. Past work has shown that people use different forms of uncertainty to guide exploration. In this study, we investigate the role of the pupil-linked arousal system in uncertainty-guided exploration. We measured participants' (n = 48) pupil dilation while they performed a two-armed bandit task. Consistent with previous work, we found that people adopted a hybrid of directed, random, and undirected exploration, which are sensitive to relative uncertainty, total uncertainty, and value difference between options, respectively. We also found a positive correlation between pupil size and total uncertainty. Furthermore, augmenting the choice model with subject-specific total uncertainty estimates decoded from the pupil size improved predictions of held-out choices, suggesting that people used the uncertainty estimate encoded in pupil size to decide which option to explore. Together, the data shed light on the computations underlying uncertainty-driven exploration. Under the assumption that pupil size reflects locus coeruleus-norepinephrine neuromodulatory activity, these results also extend the theory of the locus coeruleus-norepinephrine function in exploration, highlighting its selective role in driving uncertainty-guided random exploration. ■

INTRODUCTION

Uncertainty lies at the heart of the explore–exploit dilemma. Incomplete knowledge of the world makes the best course of action ambiguous; exploration can help reduce uncertainty by collecting information about different options, at the cost of not exploiting the option currently estimated to be the best. The computational intractability of balancing exploration and exploitation optimally has motivated the search for efficient heuristics, some of which are used by people (Schulz & Gershman, 2019). In this article, we examined how the computation underlying exploration is represented in the physiological state as assessed with pupil dilation, which has long been seen as indicative of the arousal state closely linked to learning under uncertainty (Joshi & Gold, 2020; de Berker et al., 2016; Lavin, San Martín, & Rosales Jubal, 2014; Nassar et al., 2012).

The locus coeruleus-norepinephrine (LC-NE) neuromodulatory system is one major source of the pupil-linked arousal responses, and NE has long been suggested to play an important role in controlling the balance between exploration and exploitation (Doya, 2002; Usher, Cohen, Servan-Schreiber, Rajkowski, & Aston-Jones, 1999), with supporting evidence documented in both animals and humans (Hayes & Petrov, 2016; Jepma & Nieuwenhuis, 2011; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Usher et al., 1999). Specifically, two modes of NE activity have been identified and linked to this arbitration (Aston-Jones & Cohen, 2005). A low-tonic, high-phasic NE level mode is associated with a selective increase in neuronal responsivity to task-related stimuli, which promotes exploitation. In contrast, a high-tonic, low-phasic NE level mode is associated with an overall increase in neuronal response to both task-related and non-task-related stimuli (i.e., the neuronal response is less discriminative). This neural firing pattern could promote disengagement from the current task/stimuli, leading to processing other tasks/stimuli (exploration). When assessing LC-NE activity, pupil size is often used as a non-invasive proxy of NE (Joshi & Gold, 2020). There is a consistent positive correlation between LC activation and pupil size, both during rest and when performing cognitive tasks (Reimer et al., 2016). In addition, it has been found that manipulating LC-NE level leads to pupil size change (Joshi, Li, Kalwani, & Gold, 2016), corroborating a tight link between NE level and pupil size.

One commonality across the above-mentioned studies on exploration is that they implicitly treat exploration as a unitary construct. In other words, they only distinguish between an exploitative and an exploratory state. Existing literature on computational models of exploration, however, has revealed a more complicated picture. Early research has found that people choose options proportional to their value (i.e., picking the high value option more often and exploring the low value option from time to time; see Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006). In the terminology of reinforcement learning, this value-driven exploration is captured by a softmax function,

[1]Harvard University, Cambridge, MA, [2]Center for Brains, Minds, and Machines, MIT, Cambridge, MA

matching the probability of choosing one option to the exponentiated value difference between options. Recent work has shown that people also use uncertainty to guide their exploration (Wilson, Bonawitz, Costa, & Ebitz, 2021; Gershman, 2018, 2019; Schwartenbeck et al., 2019; Wu, Schulz, Speekenbrink, Nelson, & Meder, 2018; Speekenbrink & Konstantinidis, 2015; Wilson, Geana, White, Ludvig, & Cohen, 2014; Frank, Doll, Oas-Terpstra, & Moreno, 2009). These uncertainty-driven exploration strategies can be further categorized into two families with distinct flavors. One type of uncertainty-driven exploration—directed exploration—drives exploration toward relatively uncertain options in the environment (e.g., try a new coffee shop instead of visiting your favorite one), which is equivalent to adding an uncertainty bonus to the value of the option (Dubois et al., 2021; Gershman, 2018; Wu et al., 2018). This exploration strategy is sensitive to the relative uncertainty (RU) between options, and can be formulated in reinforcement learning theories such as the Upper Confidence Bound algorithm (Auer, 2002). In contrast, random exploration injects noise into the choice process, which scales with total uncertainty (TU; e.g., explore more when moving to a new city, where there is high uncertainty across all the coffee shops). Similar to directed exploration, random exploration can also be captured by reinforcement learning algorithms such as Thompson sampling (Thompson, 1933).

Past work has identified the behavioral signatures of these two kinds of uncertainty-driven exploration by experimentally separating directed and random exploration by manipulating relative and TU, respectively (Dubois et al., 2021; Gershman, 2019; Wilson et al., 2014). Beyond the behavioral level, there is also evidence that these two uncertainty-driven exploration strategies are neurally dissociable (Tomov, Truong, Hundia, & Gershman, 2020; Gershman & Tzovaras, 2018; Zajkowski, Kossut, & Wilson, 2017), have distinct development trajectories (Meder, Wu, Schulz, & Ruggeri, 2021; Somerville et al., 2017), and are associated with trait anxiety in different ways (Fan, Gershman, & Phelps, 2023; Smith et al., 2022). Given NE's role in signaling uncertainty and guiding behavior under uncertainty (Zhao et al., 2019; Dayan, 2012; Nassar et al., 2012; Yu & Dayan, 2005), it is likely that the LC-NE system is involved in guiding uncertainty-driven exploration. However, which (if any) exploration strategy the LC-NE system drives remains unclear.

Given the role of the LC-NE system in modulating neural activity, researchers have speculated that it plays a role in driving exploration strategies that are closely related to choice randomness, which is supported by existing pharmacological studies (Cremer, Kalbe, Müller, Wiedemann, & Schwabe, 2022; Dubois et al., 2021; Warren et al., 2017). Specifically, a link between LC-NE and uncertainty-guided random exploration has been proposed (Wilson et al.,, 2014 2021; Warren et al., 2017). Increased TU could be reflected as a higher overall level of neural noise—corresponding to the tonic LC-NE mode—which could lead to more variable choices. One study dissociated different exploration strategies and examined the role of NE in uncertainty-driven exploration using atomoxetine, a NE transporter blocker that increases the NE level (Warren et al., 2017). Although they found a selective relationship between the treatment and uncertainty-guided random exploration, the direction of the link is opposite to what had been hypothesized, that is, the treatment group demonstrated less random exploration. The authors reasoned that the counterintuitive finding may be a result of atomoxetine treatment elevating phasic NE more than tonic NE. Because they did not include physiological measurement, it is hard to check the exact impact of atomoxetine on NE.

The hypothesized relationship between NE and uncertainty-guided random exploration is in line with previous work showing that pupil size—as an indirect measure of NE level (Joshi & Gold, 2020; Joshi et al., 2016)—correlates with variability in the evidence accumulation process (Cavanagh, Wiecki, Kochar, & Frank, 2014; Murphy, Vandekerckhove, & Nieuwenhuis, 2014), magnitude of noise in perceptual tasks (Keung, Hagen, & Wilson, 2019), and choice randomness in value-based decision making (Kozunova et al., 2022; Findling, Skvortsova, Dromnelle, Palminteri, & Wyart, 2019; Van Slooten, Jahfari, Knapen, & Theeuwes, 2018; Kane et al., 2017; Jepma & Nieuwenhuis, 2011). These pupillometry studies suggest that pupil size could represent the computation of TU during reinforcement learning, although this has not been directly tested yet.

In the current study, we systematically investigated the involvement of pupil-linked arousal systems in uncertainty-driven exploration by monitoring participants' pupil size while they performed a two-armed bandit task. Given previous research suggesting a relationship between tonic NE and uncertainty-guided random exploration, we focus on the pupil size before the stimulus onset, which is argued to be a readout of the tonic NE level (Joshi & Gold, 2020).

METHODS

Participants

Fifty-five healthy, neurotypical participants (age 18–65 years) were recruited from the general and student population. All participants were fluent in English and had normal-to-corrected vision and hearing. Within the past year, all participants went untreated for the following: neurological disorder or injury, psychiatric disorder, endocrine disease, blood pressure issues, and cardiac issues. Participants consumed no more than three alcoholic drinks/day on average and did not smoke cigarettes or use recreational drugs daily. Participants were requested not to consume drugs, alcohol, caffeine, or nicotine within 2 hr before completing the study. Participants were either paid $30 or received course credit. All participants

could additionally earn a performance bonus up to $5 based on their task performance. Participants completed the task in either contact lenses or glare-resistant glasses with mandatory removal of all eye makeup.

The number of participants were determined by a power analysis conducted before the data collection, which showed that given the current task setup, a sample size of around 50 (after exclusion) could reach a power of 80%. We excluded participants if they chose the more rewarding option in < 60% of all trials ($n = 1$) or did not have pupillometry data because of technique issues ($n = 7$). Forty-eight participants were included in the final data analysis (31 women, 16 men, one unreported; age $M = 32.74$ years, $SD = 12.37$ years). All participants signed an informed consent form before the experiment, and the study was approved by the Harvard University Committee on the Use of Human Subjects (IRB19-0789).

Experiment Design

Two-armed Bandit Task

Participants completed 16 blocks, 10 trials per block, of a two-armed bandit task adapted from Gershman (2019). In each block, participants encountered a new pair of slot machines and chose repeatedly between them. The mean rewards ($\mu(k)$) of both arms were randomly sampled at the start of every block from a Gaussian distribution with mean 0 and a variance ($\tau_0^2(k) = 36$) and remained the same within a block. Participants faced two types of arms: the "risky" arms, as indicated by an R, delivered variable rewards drawn from a Gaussian with mean ($\mu(R)$) and variance ($\tau^2(R) = 16$); the "safe" arms, as indicated by an S, delivered a reward of ($\mu(S)$) with no variance. On every trial, participants received feedback in the form of points, rounded to the nearby integer, from their chosen arm; no feedback was provided for missed trials. At the start of every block, participants were told the two arms they would see in the upcoming block and could see the letter R/S on the slot machines throughout the whole 10-trial block. To make sure the participants understood the task, they completed a comprehension check, performed one practice block, and had a chance to ask any question before the main experiment.

Experiment Procedure

Participants completed the task in a silent dark room, with their heads placed on a chinrest 55 cm away from the screen. The stimuli were presented on a black Dell 21.5-in. LED Monitor, and their right eye was tracked using an EyeLink 1000 plus eye tracker (SR Research) at a sample rate of 1000 Hz. The eye tracker was calibrated using a standard 9-point calibration procedure. After calibration, the participants first underwent a 5-min baseline period, during which they were asked to maintain a normal gaze at a central white fixation cross on the screen. They then completed the main experiment. A drift check (i.e., 1-point calibration) was performed at the beginning of each block to confirm that the pupil was successfully tracked. Each trial is composed of four phases (Figure 1A). In the fixation phase, a white fixation cross appears on the screen for a time period between 1 sec and 2 sec, during which the participants maintained a normal gaze at the fixation cross. The length of the fixation phase varies from trial to trial to reduce the potential impact of temporal expectation on pupil size. In the evaluation phase, the participant saw the white fixation as well as two slot machines on the screen for 3 sec. The participant was required to hold off their decision until the decision phase, which was signaled by the white fixation turning into green. The decision phase lasts 1.5 sec, during which participants could press the left or right arrow key to indicate their choice. If the participant made a decision in this 1.5-sec time window, in the feedback phase, the reward delivered by the chosen slot machine would show up inside the box representing the option and would remain on the screen for 3 sec. If no choice was made, the fixation turned red and remained for 3 sec to indicate a missed trial. The lengths of the evaluation and feedback phase follow the guideline of designing pupil experiments (i.e., the stimulus onset should be followed by an interval of 2–3 sec to ensure achieving reliable pupil measures; Mathôt & Vilotijević, 2022) and are in line with previous pupillometry research (e.g., Lempert, Lackovic, Tobe, Glimcher, & Phelps, 2017). Efforts have been made to match visual stimuli across stages and between conditions on as many low-level properties as possible, including using color of equal luminance, indicating the start of choice phase by changing the color of the fixation while keeping all other visual display else the same, displaying the feedback inside the chosen slot machine box while keeping unchosen slot machine box on the screen, as well as using the same font size to indicate different block conditions.

Belief Updating Model

Following previous literature (Fan et al., 2023; Tomov et al., 2020; Gershman, 2018, 2019; Daw et al., 2006), we assumed that participants approximate an ideal Bayesian observer, tracking the value and uncertainty in their estimation. Given the underlying Gaussian distributions of the slot machines, the posterior estimate of arm k (k = 1 denotes the left arm, k = 0 denotes the right arm) is a Gaussian distribution parameterized with mean Q(k) and variance $\sigma^2(k)$. We modeled participants' belief update process using Kalman filtering equations:

$$Q_{t+1}(a_t) = Q_t(a_t) + \alpha_t(r_t - Q_t(a_t)) \quad (1)$$

$$\sigma^2_{t+1}(a_t) = \sigma^2_t(a_t) - \alpha_t\sigma^2_t(a_t) \quad (2)$$

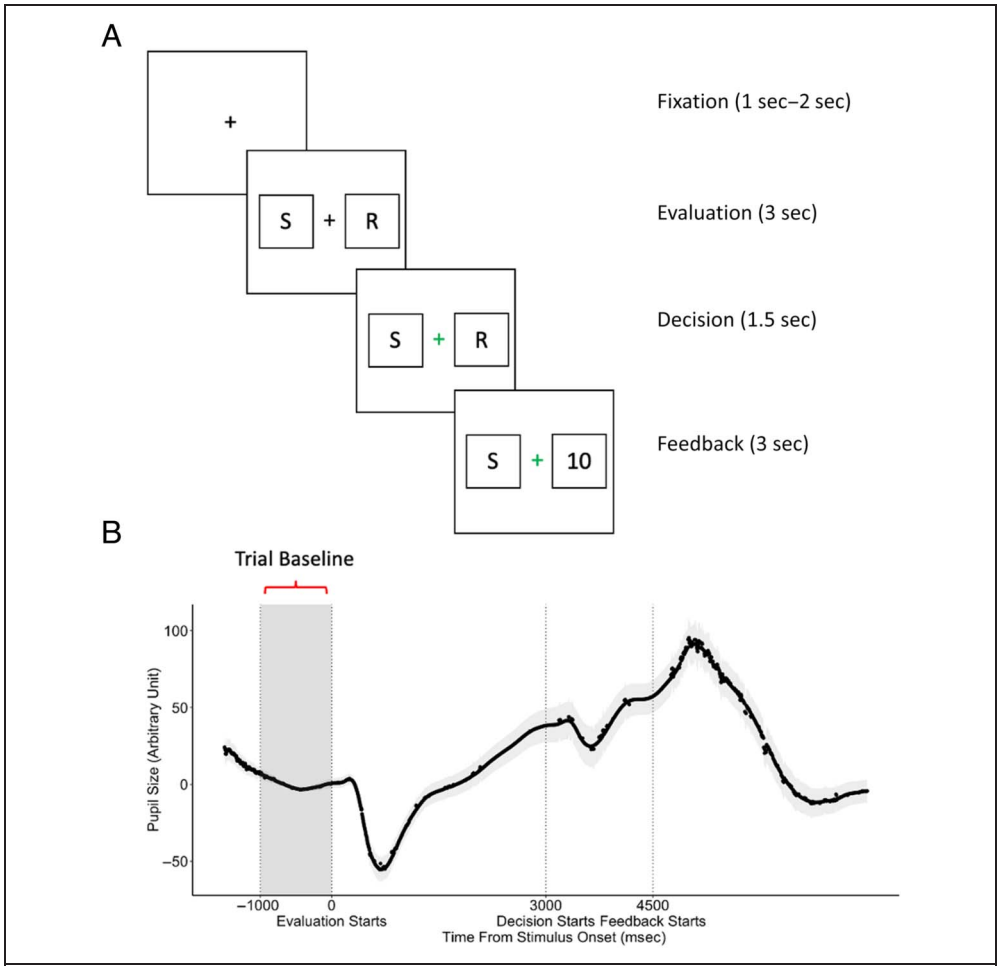

Figure 1. Trial schematic (A) and average pupil size over the course of a trial, time locked to stimulus onset (B). Each trial contains four phases: fixation, evaluation, decision, and feedback. Trial baseline pupil is defined as the pupil size averaged over the 1000 msec before the stimulus onset. The error bar represents 1 *SE*. The *y* axis shows the arbitrary unit of pupil size recorded by EyeLink 1000, which is proportional to the true diameter of the physical pupil.

where a_t is the chosen arm at time t and Kalman gain α_t is given by:

$$\alpha_t = \frac{\sigma_t^2(a_t)}{\sigma_t^2(a_t) + \tau^2(a_t)} \quad (3)$$

τ^2 was set at 16 for risky slot machines and 0.00001 for safe slot machines to avoid numerical overflow. Within a block, the learning rate for safe slots quickly approached 1 whereas the learning rate for risky slots slowly decreased.

The initial mean for risky and safe slot machines on Trial 1 was set at 0, and the initial variance for all slot machines on Trial 1 was set to 36 ($\tau_0^2 = 36$) because the participant has no prior knowledge of the reward history yet. The mean and the variance of the slot machines were only adjusted for the slot machine that the participant has chosen. Kalman filtering is an idealization of learning in a noisy environment and has been shown to account for human choice behavior in multi-armed bandit tasks (Speekenbrink & Konstantinidis, 2015; Daw et al., 2006).

Choice Probability Analysis

Participants' data were pooled across conditions and modeled using the following probit regression model:

$$P(a_t = 1|w) = \Phi(w_1 V_t + w_2 RU_t + w_3 V_t/TU_t), \quad (4)$$

where $V_t = Q_t(1) - Q_t(2)$ is the value difference between two options, $RU_t = \sigma_t(1) - \sigma_t(2)$ denotes the relative uncertainty between two options, $TU_t = \sqrt{(\sigma_t^2(1) + \sigma_t^2(2))}$ denotes the total uncertainty of the current trial, and $\Phi(\bullet)$ is the standard Gaussian cumulative distribution function. The value estimate and estimation uncertainty were obtained using Kalman filtering described in the previous section. Past work (Gershman, 2018) proved that this is the exact analytic form of a hybrid of Thompson Sampling and Upper Confidence Bound algorithms. A positive w_2 means that people add an uncertainty bonus to an option's value proportional to its RU, which directs exploration toward the option they are more uncertain about. In contrast, a positive w_3 indicates that as the TU in the environment goes up, people increase choice randomness accordingly. Specifically, if $w_2 = 0$, the model is insensitive to RU and is reduced to pure random exploration. If $w_3 = 0$, the model is insensitive to TU and is

reduced to pure directed exploration. Finally, if $w_2 = w_3 = 0$, the model is only influenced by the value of options, and the strategy is similar to softmax exploration. In line with previous work (Fan et al., 2023; Tomov et al., 2020; Gershman, 2019), we fitted a generalized mixed-effects model with a probit link function, and included fixed and random effects for all regressors (i.e., V, RU, and V/TU). We used the default prior for fixed effects defined in the R *brms* package (i.e., improper flat prior over the reals; Bürkner, 2017). The random effects were restrained to be coming from a multivariate Gaussian distribution with mean zero and unknown covariance matrix. The prior for fixed effects and constraints on random effects remain the same for the mixed-effects regression models mentioned below.

To obtain a descriptive characterization of choice behavior across conditions, we also modeled choices as a function of experimental condition (SR, RS, SS, RR; where the first letter denotes the label of the option on the left and the second letter denotes the label of the option on the right):

$$P(a_t = 1|\mathrm{w}) = \Phi\left(\sum_j w_4^j \pi_{tj} + w_5^j \pi_{tj} \mathrm{V_t}\right) \quad (5)$$

where j is the experimental condition, and $\pi_{tj} = 1$ if trial belongs to condition $j = 0$ otherwise. We refer to the w_4 terms as intercepts and the w_5 terms as slopes in the following sections. The posterior distributions of the parameters were estimated using the *brms* package in R (same for the generalized mixed regression models mentioned below), and we report the median estimate of the posterior distribution as well as the 95% highest posterior density intervals (HDIs). Parameters with 95% HDIs that did not contain zero were deemed to be statistically credible.

Pupillometry Analysis

Preprocessing of the pupil data was done following the guidelines illustrated in Kret and Sjak-Shie (2019). In line with previous literature (Leong, Dziembaj, & D'Esposito, 2021; Clewett, Gasser, & Davachi, 2020; Browning, Behrens, Jocham, O'Reilly, & Bishop, 2015), we exclude trials with > 40% missing pupil data (76 trials, 0.97% of all trials), no choice was made (121 trials, 1.54% of all trials), and the trial baseline is outside of 3 *SD*s of this person's trial baseline distribution (29 trials, 0.37% of all trials).

Our variable of interest is the average pupil size during the trial baseline (i.e., 1 sec before the stimulus onset; Figure 1B). As part of the preprocessing, we subtracted the subject-specific average trial baseline (across all trials) from their pupil response data. Two sets of linear mixed models were fit to trial baseline pupil data:

$$\begin{aligned}\text{absolute model}: \text{Pupil response}_t \sim\ & |\mathrm{V_t}| + |\mathrm{RU_t}| \\ & + |\mathrm{V_t}|/\mathrm{TU_t} + \mathrm{TU_t}, \end{aligned} \quad (6)$$

$$\begin{aligned}\text{directed model}: \text{Pupil response}_t \sim\ & \mathrm{V_{t,directed}} \\ & + \mathrm{RU_{t,directed}} + \mathrm{V_{t,directed}}/\mathrm{TU_t} \\ & + \mathrm{TU_t}, \end{aligned} \quad (7)$$

where $|\mathrm{V_t}|$ and $|\mathrm{RU_t}|$ denotes absolute value difference and RU on trial t, whereas $\mathrm{V_{t,directed}}$ and $\mathrm{RU_{t,directed}}$ denotes the value difference and RU for the option that they chose on trial t. Because TU is always non-negative, $|\mathrm{TU_t}| = \mathrm{TU_t}$. Similar to the behavioral analysis, we include fixed and random effects for all regressors. We conducted model comparison to decide which model fits the current pupil data better using the Leave-One-Out Cross Validation Information Criterion (LOOIC; Vehtari, Gelman, & Gabry, 2017; Vehtari, Simpson, Gelman, Yao, & Gabry, 2015). The model comparison was done using the *performance* package in R (Lüdecke, Ben-Shachar, Patil, Waggoner, & Makowski, 2021).

Decoding Analysis

We conducted a decoding analysis to examine whether we could decode subject-specific uncertainty estimates readout from the pupil data and use them to improve the fit of the behavioral model. We used a maximum likelihood decoder; that is, we extracted the estimate of TU by inverting the generalized mixed linear model in the absolute model (Equation 6):

$$\widehat{\mathrm{TU_t}} = \left(y_t - \sum\nolimits_{i:X_{t,i} \neq |\mathrm{TU}|} X_{t,i}\beta_i\right)/\beta_{|\mathrm{TU}|}, \quad (8)$$

We first augmented the original choice model using decoded $\widehat{\mathrm{TU_t}}$. Because TU, by definition, is non-negative, we rectified negative decoded TU to be the minimum of a trial-by-trial TU estimate in the data set. To keep the model specification consistent with the hypothesis that it represents a hybrid of exploration strategies, $\widehat{\mathrm{TU_t}}$ was entered into the model in the format of $\mathrm{V_t}/\widehat{\mathrm{TU_t}}$:

$$\begin{aligned}&\mathrm{P}(\mathrm{a_t} = 1|\mathrm{w}) \\ &\quad = \Phi\left(\mathrm{w_0} + \mathrm{w_1V_t} + \mathrm{w_2RU_t} + \mathrm{w_3V_t}/\mathrm{TU_t} + \mathrm{w_4V_t}/\widehat{\mathrm{TU_t}}\right)\end{aligned} \quad (9)$$

The augmented model was compared with a vanilla choice model, which adds an intercept to Equation 4:

$$\mathrm{P}(\mathrm{a_t} = 1|\mathrm{w}) = \Phi(\mathrm{w_0} + \mathrm{w_1V_t} + \mathrm{w_2RU_t} + \mathrm{w_3V_t}/\mathrm{TU_t}) \quad (10)$$

The intercept w_0 is included because our data suggest that simply including the intercept increases the model fit, indicating that people have a bias based on the location of the slot machines. We included it here to account for the location bias and could thus focus on the improvement of model fit resulting from adding $\mathrm{V_t}/\mathrm{TU_t}$.

Besides augmenting the choice model using decoded TU, we have also constructed a series of augmented choice models (Figure 9) using other decision variables decoded from the pupil model, that is, $\widehat{\mathrm{RU_t}}$ and $\widehat{\mathrm{V_t}}$. Because the best pupil model is the absolute

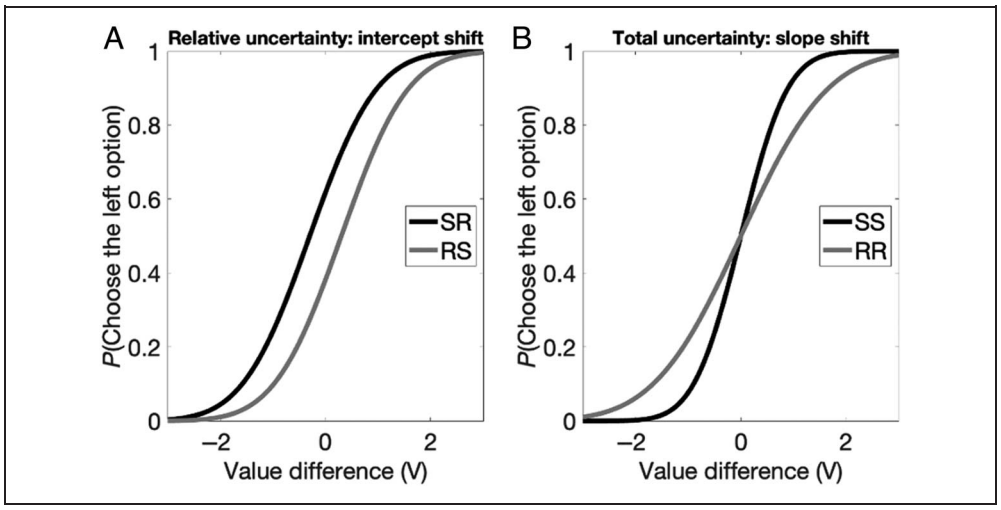

Figure 2. Predictions of choice probability function change across conditions. Directed exploration predicts a preference for the uncertain option, which manifests as a shift in intercept in opposite directions for SR and RS trials (A). Random exploration predicts more choice stochasticity when TU is high, equivalent to a steeper curve for RR trials than SS trials (B).

model, we first decode the absolute value of RU_t and V_t from Equation 6:

$$|\widehat{RU}_t| = \left(y_t - \sum_{i:X_{t,i} \neq |RU|} X_{t,i}\beta_i\right)\beta_{|RU|}/\beta^2_{|RU|} \quad (11)$$

$$|\widehat{V}_t| = \left(y_t - \sum_{i:X_{t,i} \neq |V|} X_{t,i}\beta_i\right)\beta_{|V|}/\beta^2_{|V|} \quad (12)$$

The relationship between absolute decoded value and decoded value is:

$$\begin{aligned}\widehat{RU}_t &= |\widehat{RU}_t| \text{ if } RU_t >= 0 \\ &= -|\widehat{RU}_t| \text{ if } RU_t < 0\end{aligned} \quad (13)$$

$$\begin{aligned}\widehat{V}_t &= |\widehat{V}_t| \text{ if } V_t >= 0 \\ &= -|\widehat{V}_t| \text{ if } V_t < 0\end{aligned} \quad (14)$$

Finally, $\widehat{RU}_t$ and $\widehat{V}_t$ were used to augment the choice model in a similar fashion to adding $\widehat{TU}_t$ described above. All the regressors in the decoded analysis are normalized.

RESULTS

People Use a Hybrid of Directed, Random, and Undirected Exploration Strategies

Our task design allows us to independently manipulate relative and TU so as to assess their separate influence on different uncertainty-driven exploration strategies. Because the risky arm has more variability in the delivered outcome, it is relatively uncertain compared with the safe arm. If directed exploration is sensitive to RU, we predict that people will demonstrate a preference for the risky arm when it is paired with a safe arm, that is, Option 1 in condition RS and Option 2 in condition SR. Because both conditions consist of one safe and one risky arm, the TU is held constant and the preference will be because of the change in RU. This choice bias will manifest as a larger intercept in the psychometric curves for condition RS versus SR (Figure 2A).

On the other hand, because a pair of risky arms should entail higher uncertainty than a pair of safe arms, the comparison between conditions RR and SS will indicate how TU influences exploration while controlling for RU. If random exploration, as hypothesized, scales up as TU increases, we expect that people will behave more randomly in condition RR versus SS, which will manifest as a larger slope in the psychometric curve for condition SS versus RR (Figure 2B).

In general, people performed well in the task, choosing the better option (i.e., the option that delivers more reward) 77% of the time. Consistent with our predictions, we found an intercept shift between conditions RS and SR (median estimate of the posterior $M = 0.62$, 95% HDI = [0.44, 0.79]; Figure 3A), indicating that people direct their exploration toward the option with higher RU. In addition, the intercept of RS is positive ($\Delta = 0.32$, 95%–HDI = [0.20, 0.43]) and the intercept of SR is negative ($\Delta = -0.29$, 95%–HDI = [−0.40, −0.19]), suggesting that people show a preference for the risky option regardless of its location. Together, these data suggest that people adopt a directed exploration strategy, which is sensitive to RU.

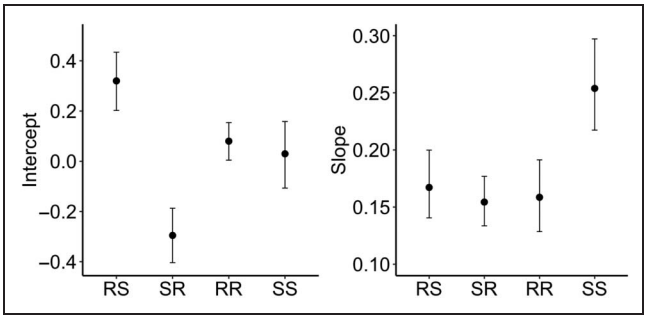

Figure 3. Probit regression results (choice data). The intercept of RS trials was larger than that of SR trials, whereas the intercepts of RR and SS trials did not differ (A). The slope of RR trials was smaller than that of SS trials, whereas the slopes of SR and RS trials did not differ (B). Error bars are 95% confidence intervals.

We also found behavioral signatures of random exploration, manifesting as a smaller slope in condition RR than SS ($\Delta = -0.1$, 95%–HDI = $[-0.15, -0.05]$; Figure 3B). In other words, in an environment with higher TU (RR), people choose both options with a probability closer to chance level (i.e., the curve is flatter). Because the slope is not different between conditions RS and SR ($\Delta = 0.01$, 95%–HDI = $[-0.02, 0.05]$), we come to the conclusion that random exploration is influenced by TU but not by RU.

Using Equation 4, we examined the exploration strategies people adopt on a trial-by-trial basis. The data suggest that people are sensitive to RU, TU, and relative value (V) when making decisions (RU: $M = 0.55$, 95%–HDI = [0.37, 0.73]; VTU: $M = 4.00$, 95%–HDI = [2.76, 5.92]; V: $M = 1.29$, 95%–HDI = [1.14, 1.48]). Taken together, our condition-based and trial-based analysis consistently show that people employ a hybrid of directed and random exploration strategies during the two-armed bandit task, which are sensitive to RU and TU, respectively.

Pupil Size Encodes TU during the Trial Baseline Period

Having demonstrated that people are sensitive to RU and TU, we examined whether these decision variables are encoded in pupil size. We used estimates of V, RU, TU, and V/TU extracted from the belief update process to predict pupil size on a trial-by-trial basis (Equations 6 and 7). Given our interest in how pupil size encodes information before the decision and the link between baseline pupil size and tonic NE, we focus our analysis on the trial

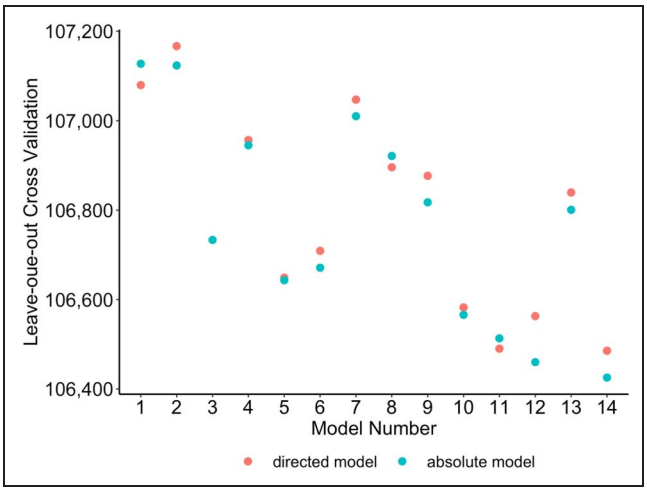

Figure 4. Model fit across nested models predicting trial baseline pupil size. *y* Axis plots LOOIC computed as −2 * expected log pointwise predictive density. Smaller LOOIC indicates better model fit. The regressors included in each models are: Model1 = V, Model2 = RU, Model3 = TU, Model4 = V/TU, Model5 = TU + RU, Model6 = TU + V, Model7 = RU + V, Model8 = V/TU + RU, Model9 = V/TU + V, Model10 = TU + V/TU, Model11 = TU + V/TU + RU, Model12 = TU + V/TU + V, Model13 = V/TU + RU + V, Model14 (full model) = V + RU + V/TU + TU.

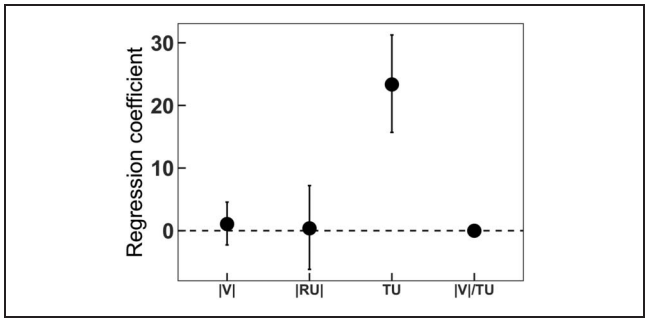

Figure 5. Coefficient estimates (extracted from the absolute model) predicting pupil size during the trial baseline. Error bars are 95% confidence intervals.

baseline. Given the block design of the task (i.e., participants interact with the same pair of slot machines within each block), pupil size during the trial baseline contains task-related information rather than mere random fluctuation. To avoid the impact of visual input change on pupil size, we calculated trial baseline pupil size using a 1-sec window before the display of the choice options (Figure 1B).

On the basis of our hypothesis that pupil size encodes the computation of uncertainty that guides exploration, we fit two models to trial baseline pupil data. The Absolute model (Equation 6) postulates that pupil dilation represents the absolute magnitude of uncertainty/value in the environment, whereas the Directed model (Equation 7) postulates that pupil size encodes the uncertainty/value estimate of the subsequent choice. Both of these models have received some empirical support in the past literature on pupil-linked arousal systems (Van Slooten et al., 2018; de Gee, Knapen, & Donner, 2014; Nassar et al., 2012). Model comparison results suggest that the absolute model outperforms the directed model (Figure 4), supporting the hypothesis that pupil size encodes the

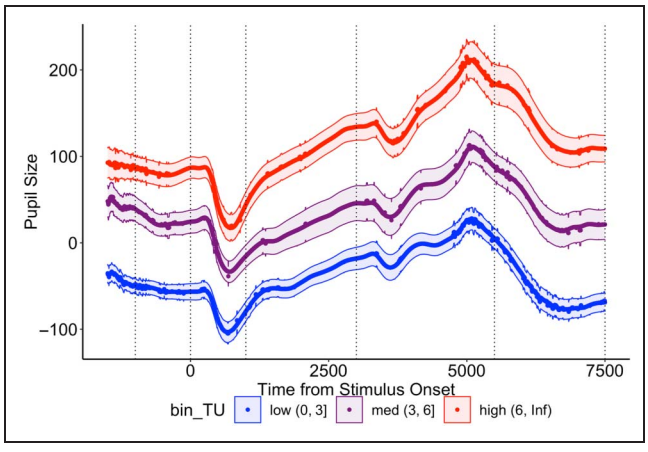

Figure 6. Average pupil diameter over the course of a trial binned by TU magnitude time locked to stimulus onset. Error bars represent standard error. The *y* axis shows the arbitrary unit of pupil size recorded by EyeLink 1000, which is proportional to the true diameter of the physical pupil.

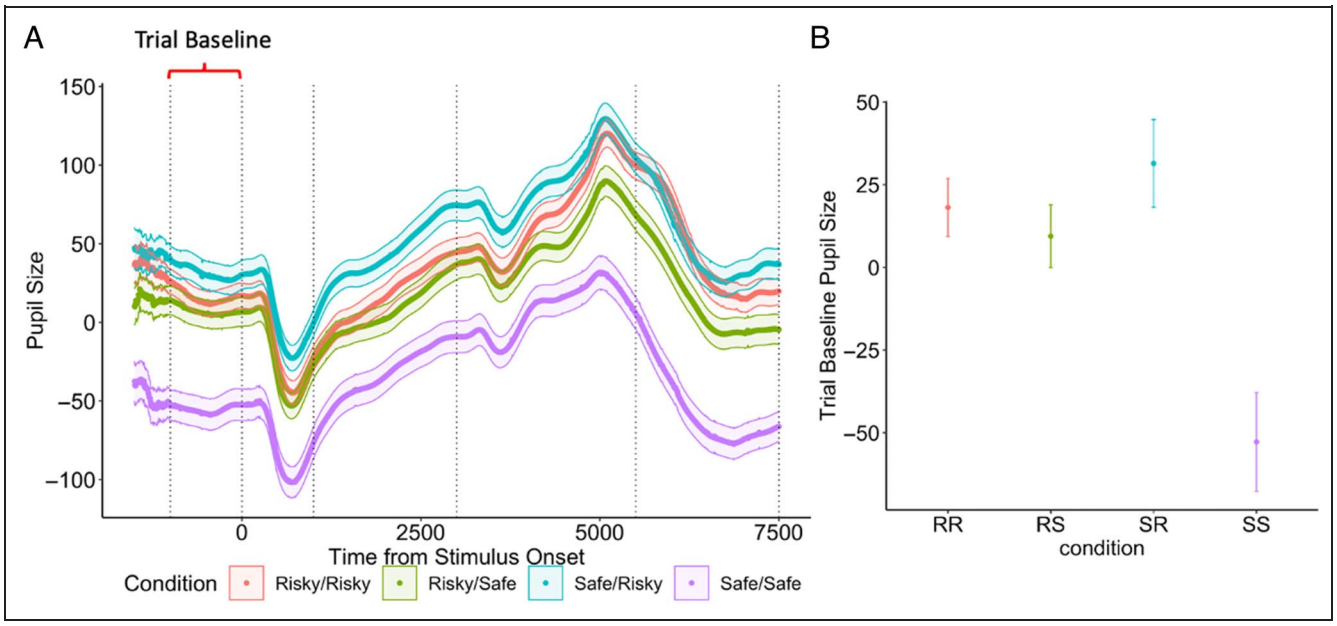

Figure 7. (A) Average pupil diameter over the course of a trial in different conditions time locked to stimulus onset and (B) average trial baseline pupil diameter (1 sec before the stimulus onset) in different conditions. Error bars represent standard error. The *y* axis shows the arbitrary unit of pupil size recorded by EyeLink 1000, which is proportional to the true diameter of the physical pupil.

absolute magnitude of the uncertainty estimates but not the directed decision variables. The absolute model suggests a positive correlation between TU and pupil size (M = 23.27, 95%–HDI = [15.69, 31.25]; Figures 5 and 6); that is, pupil size is larger on trials with higher TU. This positive association between pupil size and TU is corroborated by two more regression analyses. When we examine the average baseline pupil size in different conditions, we observe a similar effect, with pupil size smallest under SS condition, which has the lowest TU (SS vs. RR: Δ = −85.36, 95%–HDI = [−129.19, 39.69]; SS vs. SR: Δ = −62.57, 95%–HDI = [−96.49, −28.28]; SS vs. RS: Δ = −88.7, 95%–HDI = [−130.26, −47.14]; Figure 7). In addition, on a within-block level, we found that pupil size negatively correlated with the progress of the block (M = −28.91, 95%–HDI = [−36.21, −21.33]; Figure 8), where the TU gets reduced as people gain more experience interacting with the same pair of slot machine within one block. In the directed model, we observe a small positive correlation between directed RU and pupil size (M = 7.15, 95%–HDI = [2.57, 11.76]). However, because this model in general provides an inferior model fit than the absolute model (ΔLOOIC = 60.13; Figure 4), we refrain from further interpreting the results.

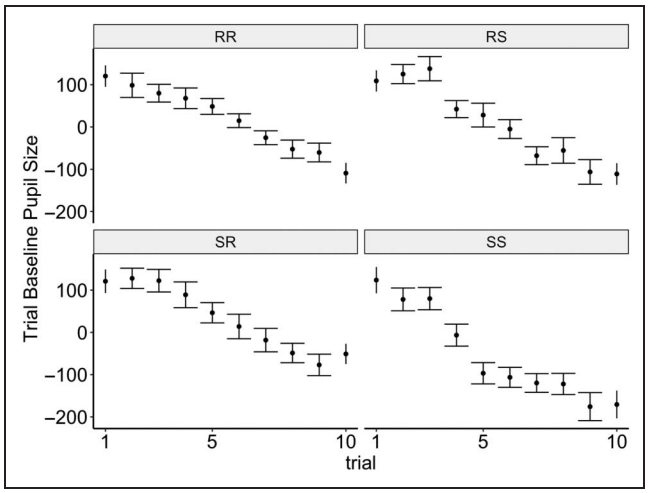

Figure 8. Average pupil diameter during trial baseline. Error bars represent standard error. The *y* axis shows the arbitrary unit of pupil size recorded by EyeLink 1000, which is proportional to the true diameter of the physical pupil.

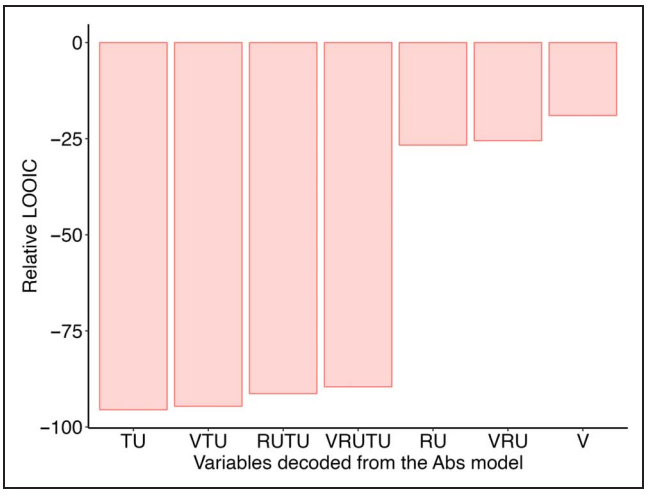

Figure 9. Model fit improvement between the augmented models and the baseline model (Equation 10). Lower relative LOOIC indicates larger improvement by adding the decoded uncertainty/value estimate. VTU = decoded V + decoded TU; VRUTU = decoded V + decoded RU + decoded TU; VTU = decoded V + decoded TU; VRU = decoded V + decoded RU.

TU Decoded from Trial Baseline Predicts Random Exploration

If pupil size encodes the TU in the environment, as our previous analysis suggests, we should be able to decode subject-specific TU estimates from the pupil size. Because the current choice model includes estimates extracted from an ideal Bayesian observer model, we reason that adding the subject-specific TU estimate would help improve the original model fitting, as we have previously demonstrated using neuroimaging data (Tomov et al., 2020). Consistent with this line of reasoning, we found that the model augmented with decoded TU derived from pupil size improved the model fit, and is the best model fit compared with other model candidates augmented with other decoded value/uncertainty estimates (ΔLOOIC = 95.50; Figure 9).

DISCUSSION

The current study examined the involvement of pupil-linked arousal systems in exploration during decision-making under uncertainty. Our analysis revealed that pupil size encodes the TU of the environment. Under the assumption that pupil size is an index of NE release from LC, this finding suggests a positive relationship between NE and uncertainty-guided random exploration, which is sensitive to TU. The link between pupil size and random exploration was corroborated by a decoding analysis, where we showed that including TU decoded from the pupil size in the choice model could improve the model fit to participants' choice. In other words, people use the TU encoded in pupil size to guide uncertainty-guided random exploration during the decision phase.

Our findings are consistent with previous work on the role of pupil-linked LC-NE system in exploration, which has shown that pupil size is larger during exploration and covaries with choice randomness (Kozunova et al., 2022; Dubois et al., 2021; Muller, Mars, Behrens, & O'Reilly, 2019; Van Slooten et al., 2018; Jepma & Nieuwenhuis, 2011). Using a two-armed bandit task, we manipulated different kinds of uncertainty and dissociated directed, random, and undirected exploration, which are sensitive to RU, TU, and V, respectively. Our data suggest that pupil size encodes TU but not RU, indicating a selective association between pupil size and uncertainty-guided random exploration. Given the covariation between pupil size and LC-NE activation, this pupillary result extended the notion that NE controls the transition between exploitation and exploration and is in line with the recent argument that the LC-NE system dynamically shapes the excitability and receptivity of neurons across the brain (Wainstein, Müller, Taylor, Munn, & Shine, 2022).

Among the three exploration strategies, uncertainty-guided random and undirected exploration both inject randomness into the choice process but differ in the source of the randomness. Undirected exploration adds a fixed level of decision noise into value computation whereas uncertainty-guided random exploration scales up choice stochasticity when TU is high (Wilson et al., 2021; Schulz & Gershman, 2019). Previous work that treats exploration as a unitary construct usually defines exploration as softmax exploration, that is, the tendency to choose the lower value option and is independent of uncertainty (Daw et al., 2006). It is often operationalized as the inverse temperature in standard softmax policy in reinforcement learning models. Softmax exploration is similar to undirected exploration in our framework, the computation of which is not reflected by pupil size. This lack of relationship is in line with a previous pharmacological study that did not find a change in softmax exploration when NE is manipulated by reboxetine (Jepma, te Beek, Wagenmakers, van Gerven, & Nieuwenhuis, 2010). However, note that because this study did not address uncertainty-driven exploration, the fitted softmax exploration parameter may reflect a combination of multiple exploration strategies and it may not be fair to directly compare to undirected exploration defined in our task.

A few recent studies have investigated the relationship between NE and different exploration strategies, most of which directly use pharmacological manipulation to change NE level and have yielded mixed results. One experiment used the horizon task (for a description of the task, see Wilson et al., 2014) and found that increasing NE level reduces uncertainty-guided random exploration (Warren et al., 2017). This positive link is further supported by another study using a sequential decision-making task (Cremer et al., 2022). On the other hand, a recent study using a task similar to the horizon task found that reducing NE level lowers the level of choice randomness (Dubois et al., 2021). Under the assumption that pupil size serves as an indirect measure of LC-NE activity, our results suggest a positive link between NE and uncertainty-guided random exploration (i.e., larger pupil size is associated with more choice stochasticity when TU is higher).

The current study focuses on pupil size during the trial baseline, which is assumed to be a proxy of tonic NE level (Joshi & Gold, 2020). Given that tonic and phasic NE have been hypothesized to play opposite roles in influencing exploration (Aston-Jones & Cohen, 2005), one potential reason for the past mixed findings could be that different pharmacological manipulations have influenced tonic and phasic NE differently. It is tricky to monitor the NE level during human pharmacological studies because different drugs may have different impacts on pupil size (Koudas et al., 2009). For example, propranolol, a commonly used drug to reduce the NE level, has been shown to not influence overall pupil size (Hauser, Eldar, & Dolan, 2017; Lempert et al., 2017) but could potentially influence how strongly pupil size encodes decision variables (Lawson, Mathys, & Rees, 2017; Lempert et al., 2017). Therefore, more research is needed to understand the impact of different pharmacological manipulations on tonic versus phasic NE.

It is important to note that the interpretation of pupil size as a real-time read-out of NE level is still under debate, with some work showing that there exists considerable variability in the strength of the coupling between pupil size and real-time NE level (Megemont, McBurney-Lin, & Yang, 2022). Another work has proposed to take both pupil size and saccade into consideration to infer the underlying neural activity (Burlingham, Mirbagheri, & Heeger, 2022). Therefore, we should interpret the pupillometry results with caution, and future studies could use more advanced analysis methods to obtain a cleaner measurement of NE from the eye-tracking data.

Our findings—again, under the assumption that pupil size can be interpreted as a readout of NE level—are also in line with previous work showing that NE covaries with learning rate (Lawson, Bisby, Nord, Burgess, & Rees, 2021; Jepma et al., 2018; Nassar et al., 2012), because normative accounts of learning have suggested that people should update the action-reward contingency more quickly when TU is high (Piray & Daw, 2021; Pulcu & Browning, 2019; Marshall et al., 2016; Browning et al., 2015; Behrens, Woolrich, Walton, & Rushworth, 2007). The involvement of pupil-linked NE signals in regulating learning dynamics has been confirmed by one past study showing that a task-independent manipulation of pupil size (i.e., play a novel sound) could alter the degree of which people learn from the new incoming data (Nassar et al., 2012). Combined with our finding that we are able to use pupil data to improve the fit of the choice model, TU estimates from pupil size, it would be interesting to investigate whether a similar exogenous pupil manipulation could influence the usage of uncertainty-guided random exploration strategy.

Acknowledgments

We thank Jasmine Pan for helpful discussion on the experiment design and data analysis.

Reprint requests should be sent to Haoxue Fan, Department of Psychology, Harvard University, 52 Oxford Street, Cambridge, MA, 02138, or via e-mail: haoxue_fan@g.harvard.edu.

Data Availability Statement

All code is publicly available at https://github.com/haoxue-fan/pupilexplore, including code related to running the experiment (instructions, comprehension check, and task), model simulation, data analysis, and figure creation. The de-identified data are publicly available at https://osf.io/q795b/?view_only=1a86356258534a5498904a393dbe0692.

Author Contributions

Haoxue Fan: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Validation; Visualization; Writing—Original draft Preparation; Writing—Review & editing. Taylor Burke: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Visualization; Writing—Original draft; Writing—Review & editing. Deshawn Chatman Sambrano: Formal analysis; Software. Emily Dial: Data curation; Investigation; Software; Validation; Visualization; Writing—Original draft; Writing—Review & editing. Elizabeth A. Phelps: Conceptualization; Funding acquisition; Methodology; Project administration; Resources; Supervision; Writing—Review & editing. Samuel J. Gershman: Conceptualization; Funding acquisition; Methodology; Supervision; Writing—Review & editing.

Funding Information

This work was supported by the National Institute on Drug Abuse (grant no. R01 DA042855 to E. A. P.) and the Air Force Office of Scientific Research (grant no. FA9550-22-1-0345). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

REFERENCES

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709, PubMed: 16022602

Auer, P. (2002). Using confidence bounds for exploitation–exploration trade-offs. *Journal of Machine Learning Research*, *3*, 397–422.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*, 1214–1221. https://doi.org/10.1038/nn1954, PubMed: 17676057

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*, 590–596. https://doi.org/10.1038/nn.3961, PubMed: 25730669

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Burlingham, C. S., Mirbagheri, S., & Heeger, D. J. (2022). A unified model of the task-evoked pupil response. *Science Advances*, *8*, eabi9979. https://doi.org/10.1126/sciadv.abi9979, PubMed: 35442730

Cavanagh, J. F., Wiecki, T. V., Kochar, A., & Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology: General*, *143*, 1476–1488. https://doi.org/10.1037/a0035813, PubMed: 24548281

Clewett, D., Gasser, C., & Davachi, L. (2020). Pupil-linked arousal signals track the temporal organization of events in memory. *Nature Communications*, *11*, 4007. https://doi.org/10.1038/s41467-020-17851-9, PubMed: 32782282

Cremer, A., Kalbe, F., Müller, J. C., Wiedemann, K., & Schwabe, L. (2022). Disentangling the roles of dopamine and noradrenaline in the exploration-exploitation tradeoff during human decision-making. *Neuropsychopharmacology*, *48*, 1078–1086. https://doi.org/10.1038/s41386-022-01517-9, PubMed: 36522404

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*, 876–879. https://doi.org/10.1038/nature04766, PubMed: 16778890

Dayan, P. (2012). Twenty-five lessons from computational neuromodulation. *Neuron*, *76*, 240–256. https://doi.org/10.1016/j.neuron.2012.09.027, PubMed: 23040818

de Berker, A. O., Rutledge, R. B., Mathys, C., Marshall, L., Cross, G. F., Dolan, R. J., et al. (2016). Computations of uncertainty mediate acute stress responses in humans. *Nature Communications*, *7*, 10996. https://doi.org/10.1038/ncomms10996, PubMed: 27020312

de Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences, U.S.A.*, *111*, E618–E625. https://doi.org/10.1073/pnas.1317557111, PubMed: 24449874

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, *15*, 495–506. https://doi.org/10.1016/S0893-6080(02)00044-8, PubMed: 12371507

Dubois, M., Habicht, J., Michely, J., Moran, R., Dolan, R. J., & Hauser, T. U. (2021). Human complex exploration strategies are enriched by noradrenaline-modulated heuristics. *eLife*, *10*, e59907. https://doi.org/10.7554/eLife.59907, PubMed: 33393461

Fan, H., Gershman, S. J., & Phelps, E. A. (2023). Trait somatic anxiety is associated with reduced directed exploration and underestimation of uncertainty. *Nature Human Behaviour*, *7*, 102–113. https://doi.org/10.1038/s41562-022-01455-y, PubMed: 36192493

Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S., & Wyart, V. (2019). Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nature Neuroscience*, *22*, 2066–2077. https://doi.org/10.1038/s41593-019-0518-9, PubMed: 31659343

Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience*, *12*, 1062–1068. https://doi.org/10.1038/nn.2342, PubMed: 19620978

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, *173*, 34–42. https://doi.org/10.1016/j.cognition.2017.12.014, PubMed: 29289795

Gershman, S. J. (2019). Uncertainty and exploration. *Decision*, *6*, 277–286. https://doi.org/10.1037/dec0000101, PubMed: 33768122

Gershman, S. J., & Tzovaras, B. G. (2018). Dopaminergic genes are associated with both directed and random exploration. *Neuropsychologia*, *120*, 97–104. https://doi.org/10.1016/j.neuropsychologia.2018.10.009, PubMed: 30347192

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, *10*, 252–269. https://doi.org/10.3758/CABN.10.2.252, PubMed: 20498349

Hauser, T. U., Eldar, E., & Dolan, R. J. (2017). Separate mesocortical and mesolimbic pathways encode effort and reward learning signals. *Proceedings of the National Academy of Sciences, U.S.A.*, *114*, E7395–E7404. https://doi.org/10.1073/pnas.1705643114, PubMed: 28808037

Hayes, T. R., & Petrov, A. A. (2016). Mapping and correcting the influence of gaze position on pupil size measurements. *Behavior Research Methods*, *48*, 510–527. https://doi.org/10.3758/s13428-015-0588-x, PubMed: 25953668

Jepma, M., Brown, S. B. R. E., Murphy, P. R., Koelewijn, S. C., de Vries, B., van den Maagdenberg, A. M., et al. (2018). Noradrenergic and cholinergic modulation of belief updating. *Journal of Cognitive Neuroscience*, *30*, 1803–1820. https://doi.org/10.1162/jocn_a_01317, PubMed: 30063180

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration–exploitation trade-off: Evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, *23*, 1587–1596. https://doi.org/10.1162/jocn.2010.21548, PubMed: 20666595

Jepma, M., te Beek, E. T., Wagenmakers, E.-J., van Gerven, J. M. A., & Nieuwenhuis, S. (2010). The role of the noradrenergic system in the exploration–exploitation trade-off: A psychopharmacological study. *Frontiers in Human Neuroscience*, *4*, 170. https://doi.org/10.3389/fnhum.2010.00170, PubMed: 21206527

Joshi, S., & Gold, J. I. (2020). Pupil size as a window on neural substrates of cognition. *Trends in Cognitive Sciences*, *24*, 466–480. https://doi.org/10.1016/j.tics.2020.03.005, PubMed: 32331857

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*, 221–234. https://doi.org/10.1016/j.neuron.2015.11.028, PubMed: 26711118

Kane, G. A., Vazey, E. M., Wilson, R. C., Shenhav, A., Daw, N. D., Aston-Jones, G., et al. (2017). Increased locus coeruleus tonic activity causes disengagement from a patch-foraging task. *Cognitive, Affective, & Behavioral Neuroscience*, *17*, 1073–1083. https://doi.org/10.3758/s13415-017-0531-y, PubMed: 28900892

Keung, W., Hagen, T. A., & Wilson, R. C. (2019). Regulation of evidence accumulation by pupil-linked arousal processes. *Nature Human Behaviour*, *3*, 636–645. https://doi.org/10.1038/s41562-019-0551-4, PubMed: 31190022

Koudas, V., Nikolaou, A., Hourdaki, E., Giakoumaki, S. G., Roussos, P., & Bitsios, P. (2009). Comparison of ketanserin, buspirone and propranolol on arousal, pupil size and autonomic function in healthy volunteers. *Psychopharmacology*, *205*, 1–9. https://doi.org/10.1007/s00213-009-1508-5, PubMed: 19288084

Kozunova, G. L., Sayfulina, K. E., Prokofyev, A. O., Medvedev, V. A., Rytikova, A. M., Stroganova, T. A., et al. (2022). Pupil dilation and response slowing distinguish deliberate explorative choices in the probabilistic learning task. *Cognitive, Affective, & Behavioral Neuroscience*, *22*, 1108–1129. https://doi.org/10.3758/s13415-022-00996-z, PubMed: 35359274

Kret, M. E., & Sjak-Shie, E. E. (2019). Preprocessing pupil size data: Guidelines and code. *Behavior Research Methods*, *51*, 1336–1342. https://doi.org/10.3758/s13428-018-1075-y, PubMed: 29992408

Lavin, C., San Martín, R., & Rosales Jubal, E. (2014). Pupil dilation signals uncertainty and surprise in a learning gambling task. *Frontiers in Behavioral Neuroscience*, *7*, 218. https://doi.org/10.3389/fnbeh.2013.00218, PubMed: 24427126

Lawson, R. P., Bisby, J., Nord, C. L., Burgess, N., & Rees, G. (2021). The computational, pharmacological, and physiological determinants of sensory learning under uncertainty. *Current Biology*, *31*, 163–172. https://doi.org/10.1016/j.cub.2020.10.043, PubMed: 33188745

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience*, *20*, 1293–1299. https://doi.org/10.1038/nn.4615, PubMed: 28758996

Lempert, K. M., Lackovic, S. F., Tobe, R. H., Glimcher, P. W., & Phelps, E. A. (2017). Propranolol reduces reference-dependence in intertemporal choice. *Social Cognitive and Affective Neuroscience*, *12*, 1394–1401. https://doi.org/10.1093/scan/nsx081, PubMed: 28992268

Leong, Y. C., Dziembaj, R., & D'Esposito, M. (2021). Pupil-linked arousal biases evidence accumulation toward desirable percepts during perceptual decision-making. *Psychological Science*, *32*, 1494–1509. https://doi.org/10.1177/09567976211004547, PubMed: 34473581

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*, 3139. https://doi.org/10.21105/joss.03139

Marshall, L., Mathys, C., Ruge, D., de Berker, A. O., Dayan, P., Stephan, K. E., et al. (2016). Pharmacological fingerprints of contextual uncertainty. *PLoS Biology*, *14*, e1002575. https://doi.org/10.1371/journal.pbio.1002575, PubMed: 27846219

Mathôt, S., & Vilotijević, A. (2022). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods*, 1–23. https://doi.org/10.3758/s13428-022-01957-7, PubMed: 36028608

Meder, B., Wu, C. M., Schulz, E., & Ruggeri, A. (2021). Development of directed and random exploration in children. *Developmental Science*, *24*, e13095. https://doi.org/10.1111/desc.13095, PubMed: 33539647

Megemont, M., McBurney-Lin, J., & Yang, H. (2022). Pupil diameter is not an accurate real-time readout of locus coeruleus activity. *eLife*, *11*, e70510. https://doi.org/10.7554/eLife.70510, PubMed: 35107419

Muller, T. H., Mars, R. B., Behrens, T. E., & O'Reilly, J. X. (2019). Control of entropy in neural models of environmental state. *eLife*, *8*, e39404. https://doi.org/10.7554/eLife.39404, PubMed: 30816090

Murphy, P. R., Vandekerckhove, J., & Nieuwenhuis, S. (2014). Pupil-linked arousal determines variability in perceptual decision making. *PLoS Computational Biology*, *10*, e1003854. https://doi.org/10.1371/journal.pcbi.1003854, PubMed: 25232732

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*, 1040–1046. https://doi.org/10.1038/nn.3130, PubMed: 22660479

Piray, P., & Daw, N. D. (2021). A model for learning based on the joint estimation of stochasticity and volatility. *Nature Communications*, *12*, 6587. https://doi.org/10.1038/s41467-021-26731-9, PubMed: 34782597

Pulcu, E., & Browning, M. (2019). The misestimation of uncertainty in affective disorders. *Trends in Cognitive Sciences*, *23*, 865–875. https://doi.org/10.1016/j.tics.2019.07.007, PubMed: 31431340

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., et al. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, *7*, 13289. https://doi.org/10.1038/ncomms13289, PubMed: 27824036

Schulz, E., & Gershman, S. J. (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, PubMed: 30529148

Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H., Kronbichler, M., & Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *eLife*, *8*, e41703. https://doi.org/10.7554/eLife.41703, PubMed: 31074743

Smith, R., Taylor, S., Wilson, R. C., Chuning, A. E., Persich, M. R., Wang, S., et al. (2022). Lower levels of directed exploration and reflective thinking are associated with greater anxiety and depression. *Frontiers in Psychiatry*, *12*, 782136. https://doi.org/10.3389/fpsyt.2021.782136, PubMed: 35126200

Somerville, L. H., Sasse, S. F., Garrad, M. C., Drysdale, A. T., Abi Akar, N., Insel, C., et al. (2017). Charting the expansion of strategic exploratory behavior during adolescence. *Journal of Experimental Psychology: General*, *146*, 155–164. https://doi.org/10.1037/xge0000250, PubMed: 27977227

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, PubMed: 25899069

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, *25*, 285–294. https://doi.org/10.1093/biomet/25.3-4.285

Tomov, M. S., Truong, V. Q., Hundia, R. A., & Gershman, S. J. (2020). Dissociable neural correlates of uncertainty underlie different exploration strategies. *Nature Communications*, *11*, 2371. https://doi.org/10.1038/s41467-020-15766-z, PubMed: 32398675

Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowski, J., & Aston-Jones, G. (1999). The role of locus coeruleus in the regulation of cognitive performance. *Science*, *283*, 549–554. https://doi.org/10.1126/science.283.5401.549, PubMed: 9915705

Van Slooten, J. C., Jahfari, S., Knapen, T., & Theeuwes, J. (2018). How pupil responses track value-based decision-making during and after reinforcement learning. *PLoS Computational Biology*, *14*, e1006632. https://doi.org/10.1371/journal.pcbi.1006632, PubMed: 30500813

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

Vehtari, A., Simpson, D., Gelman, A., Yao, Y., & Gabry, J. (2015). Pareto smoothed importance sampling. *arXiv:1507.02646*. https://doi.org/10.48550/arXiv.1507.02646

Wainstein, G., Müller, E. J., Taylor, N., Munn, B., & Shine, J. M. (2022). The role of the locus coeruleus in shaping adaptive cortical melodies. *Trends in Cognitive Sciences*, *26*, 527–538. https://doi.org/10.1016/j.tics.2022.03.006, PubMed: 35469726

Warren, C. M., Wilson, R. C., van der Wee, N. J., Giltay, E. J., van Noorden, M. S., Cohen, J. D., et al. (2017). The effect of atomoxetine on random and directed exploration in humans. *PLoS One*, *12*, e0176034. https://doi.org/10.1371/journal.pone.0176034, PubMed: 28445519

Wilson, R. C., Bonawitz, E., Costa, V. D., & Ebitz, R. B. (2021). Balancing exploration and exploitation with information and randomization. *Current Opinion in Behavioral Sciences*, *38*, 49–56. https://doi.org/10.1016/j.cobeha.2020.10.001, PubMed: 33184605

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

Psychology: General, *143*, 2074–2081. https://doi.org/10.1037/a0038199, PubMed: 25347535
Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2018). Generalization guides human exploration in vast decision spaces. *Nature Human Behaviour*, *2*, 915–924. https://doi.org/10.1038/s41562-018-0467-4, PubMed: 30988442
Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*, 681–692. https://doi.org/10.1016/j.neuron.2005.04.026, PubMed: 15944135
Zajkowski, W. K., Kossut, M., & Wilson, R. C. (2017). A causal role for right frontopolar cortex in directed, but not random, exploration. *eLife*, *6*, e27430. https://doi.org/10.7554/eLife.27430, PubMed: 28914605
Zhao, S., Chait, M., Dick, F., Dayan, P., Furukawa, S., & Liao, H.-I. (2019). Pupil-linked phasic arousal evoked by violation but not emergence of regularity within rapid sound sequences. *Nature Communications*, *10*, 4030. https://doi.org/10.1038/s41467-019-12048-1, PubMed: 31492881