Pupil size reflects the relevance of reward prediction error and estimation uncertainty in upcoming choice

Zoe W. He[1*], Maëva L'Hôtellier[2], Alexander Paunov[2], Dalin Guo[1], Florent Meyniel[2], Angela J. Yu[3,4*]
[1]Department of Cognitive Science, University of California San Diego, La Jolla, California, USA
[2]INSERM-CEA Cognitive Neuroimaging Unit (UNICOG), NeuroSpin Center, CEA Paris-Saclay, Gif-sur-Yvette, France Université de Paris, Paris, France
[3]Centre for Cognitive Science & Hessian AI Centre, Technical University of Darmstadt, Darmstadt, Germany
[4]Halicioglu Data Science Institute, University of California San Diego, La Jolla, California, USA
([*]wah016@ucsd.edu, angela@angelayu.org)

Abstract

How humans process and utilize experienced outcomes and actions to adapt to a constantly evolving and noisy world is an important area of research. We investigate the role of the pupil-linked arousal system in adaptive value-based decision-making in an uncertain and changing environment using a two-armed bandit task with occasional changes in reward contingencies. We find that pupil size fluctuation encodes reward- and uncertainty-related values across trials; moreover, pupil size reflects future-choice-dependent contributions of these variables to learning and decision-making: larger pupil encoding of reward prediction error (RPE) promotes reward-driven switches in choice, while larger pupil encodings of estimation uncertainty (EU) promotes uncertainty-driven switches in choice. Furthermore, individual differences in pupil's encoding of RPE and EU correlate with individual variabilities in choice bias and task performance. Given the relationship of pupil size to noradrenergic and cholinergic modulations, these results provide insights into the computational and neural process underlying adaptive decision-making.

Keywords: decision-making; uncertainty; multi-armed bandit; Bayesian modeling; pupillometry

Introduction

Decision-making under uncertainty is a central aspect of human cognition, fundamental to navigating the complexities of daily life. In a constantly evolving and largely unpredictable world, the ability to make successful decisions hinges on our capacity to continually learn from our environment and adapt our choices accordingly. A central question in this dynamic decision-making process is understanding how experienced outcomes and previous actions guide these adaptive decisions. This study aims to examine the mechanisms and computations of adaptive choices in humans, particularly examining how they are represented in physiological states, as assessed by fluctuating pupil size, known to be related to both noradrenergic and cholinergic activity in the cortex (Joshi & Gold, 2020; Mathôt, 2018; Reimer et al., 2016; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Aston-jones & Cohen, 2005).

Despite the established links between adaptive choices, learning, and pupil response, systematic explanations of how learned factors and mechanisms influence adaptive choices are still lacking. This gap is partly due to the constrained focus of the existing studies. Many studies focus solely on learning or inference without any decision-making components (Nassar et al., 2012), or are based on perceptual decision-making that do not rely on previously learned, reward-based information (Kucewicz et al., 2018; van der Wel & van Steenbergen, 2018; Colizoli, de Gee, Urai, & Donner, 2018; Urai, Braun, & Donner, 2017). Studies on value-based decision-making tend to adopt a more stationary environment (Fan et al., 2023) or do not balance out the anti-correlation between reward and uncertainty (Jepma & Nieuwenhuis, 2011; Preuschoff, 't Hart, & Einhäuser, 2011).

To address these limitations and deepen our understanding of how the pupil-linked arousal system influences the integration of learning with adaptive choices, this study investigates pupil fluctuations in individuals engaged in a dynamic two-armed bandit task, a commonly used value-based decision-making paradigm in both behavioral sciences and machine learning that captures the interplay between learning and decision-making (Wilson, Geana, White, Ludvig, & Cohen, 2014; Cohen, McClure, & Yu, 2007). To create a changing environment, we modify the classical bandit task to have occasional "change-points" with changes in reward contingencies that influence the participants' estimation of reward and uncertainty, thereby encouraging them to modify their existing decision strategies from time to time. By analyzing the relationship between trial-to-trial updates of learned variables, choice alterations, and pupil responses, we aim to discern whether the arousal-indexed pupil response merely tracks variables such as reward prediction error (RPE) and uncertainty, or reflects a more profound relationship between learning and decision-making. This study seeks to shed light on the process by which learned variables are incorporated into actual decision-making, thereby offering new insights into the complex mechanisms of adaptive choice behavior.

Results

Post-change-point pupil fluctuation captures the variability in switch delay due to different noise levels

We examine pupil diameters from fifty-four participants playing a novel two-armed bandit decision task that involved repeated choices among two options (Figure 1A; also see Methods). On each trial, the reward outcome for each option was sampled from a Gaussian distribution with a mean that switched among three values (30, 50, and 70) at unannounced change-points throughout the task (Figure 1B). Change-points occurred independently for the two options at

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

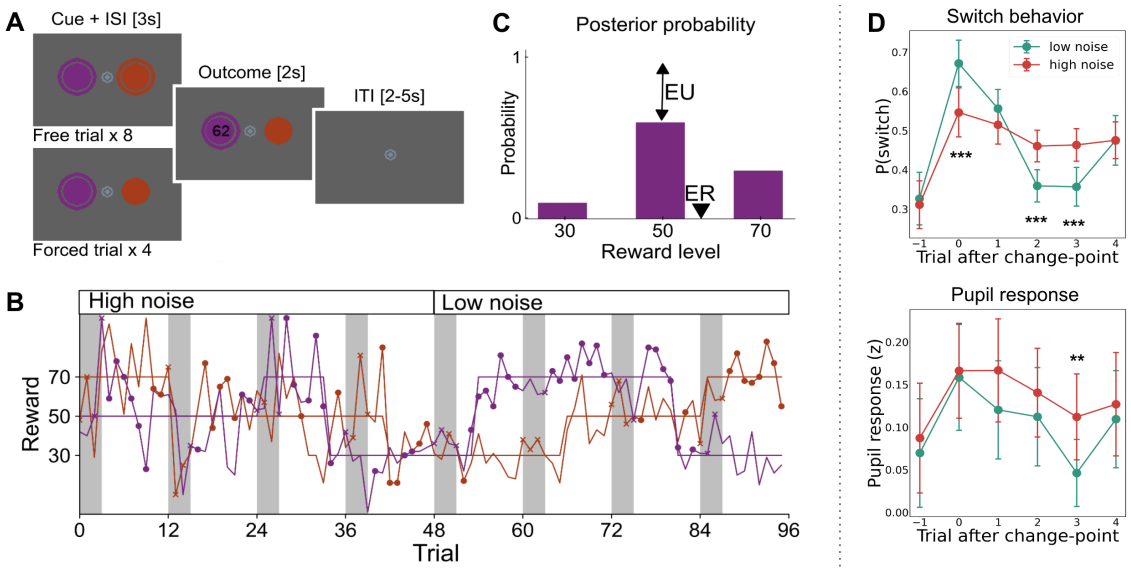

Figure 1: **A.** Task illustration. On each free trial, subjects chose between two options (purple or orange) and observed the outcome of the selected option. On each forced trial, subjects can only select the option with a circle around (purple here). **B.** An example block (96 trials) with reward generation process visualized. Solid lines: underlying mean reward levels (30, 50, or 70) for each option. Change-points are represented by abrupt shifts of reward levels. Dots: reward outcomes (between 1 to 100) drawn from the Gaussian distribution based on the mean and the noise level (high noise = 20, low noise = 10). Shaded areas: forced trials. **C.** An example of the ideal Bayesian posterior probability distribution of the reward level of an option. ER: Expected reward (posterior mean). EU: Expected uncertainty (1 minus the probability of the maximum a posteriori (MAP) reward level). **D.** Tendency to switch option (top) and outcome-evoked baselined pupil response (bottom) after a high-to-low change-point under two different noise levels.

random intervals during a game; on average, they occurred once every 24 trials (true volatility = 1/24). The standard deviation of the Gaussian reward generation process for both options changed simultaneously between two noise levels (10 or 20). The participants were explicitly informed about the current noise level in every trial.

To examine how pupil fluctuations relate to the processes of belief updating and strategy changing throughout the game, we use a Bayesian ideal observer (IO) model to compute each subject's posterior estimates of the per-trial reward level for each option given the sequence of behavioral choices and observations (see Methods). Because the Bayesian model places a distribution over the posterior reward estimates, it naturally captures the uncertainty inherent in the observation possibly represented by the brain. Using the Bayesian IO model, we derive the trial-wise estimates of expected reward (ER, defined as the weighted average of the reward levels based on posterior probability) and estimation uncertainty (EU, defined as the probability that the most likely reward level is not the actual generative reward mean; see Figure 1C).

We find that the fluctuation of pupil response across trials encodes the delay in switch actions due to different noise levels, particularly after the subjects observe a high-to-low change-point (Figure 1D). Behaviorally, subjects switch between options soon after observing a high-to-low change point on a previously preferred option and do so sooner in the (easier) low-noise condition (Figure 1D, top). The magnitude of the outcome-evoked pupil dilation also peaks and then decays shortly after the high-to-low change point, with the peak occurring earlier in the low-noise condition (Fig. 1D, bottom). This suggests that pupil response is sensitive to outcome changes, in a way that reflects the variability in adaptive choice behaviors.

Pupil response encodes the interaction between RPE and choice

We categorize each choice on a trial based on the sign of the model-fitted reward prediction error (RPE), and examine the corresponding post-outcome pupil size prior to choice. While the pupil response does not differ significantly between negative RPE ("(-)RPE") and positive RPE ("(+)RPE") conditions, significant variability emerges when the conditions are further categorized based on whether the choice on the next trial is a "switch" (shift away from the current option) or "stay" (continue to choose the current option) (Figure 2A, B). This implies that post-outcome pupil response may encode an interaction between RPE and upcoming choice, since pupil responded to RPE differently depending on the upcoming choice.

We construct a logistic regression model to predict the next choice action using RPE, pupil response, and their interaction terms. We compare the performance of the model with

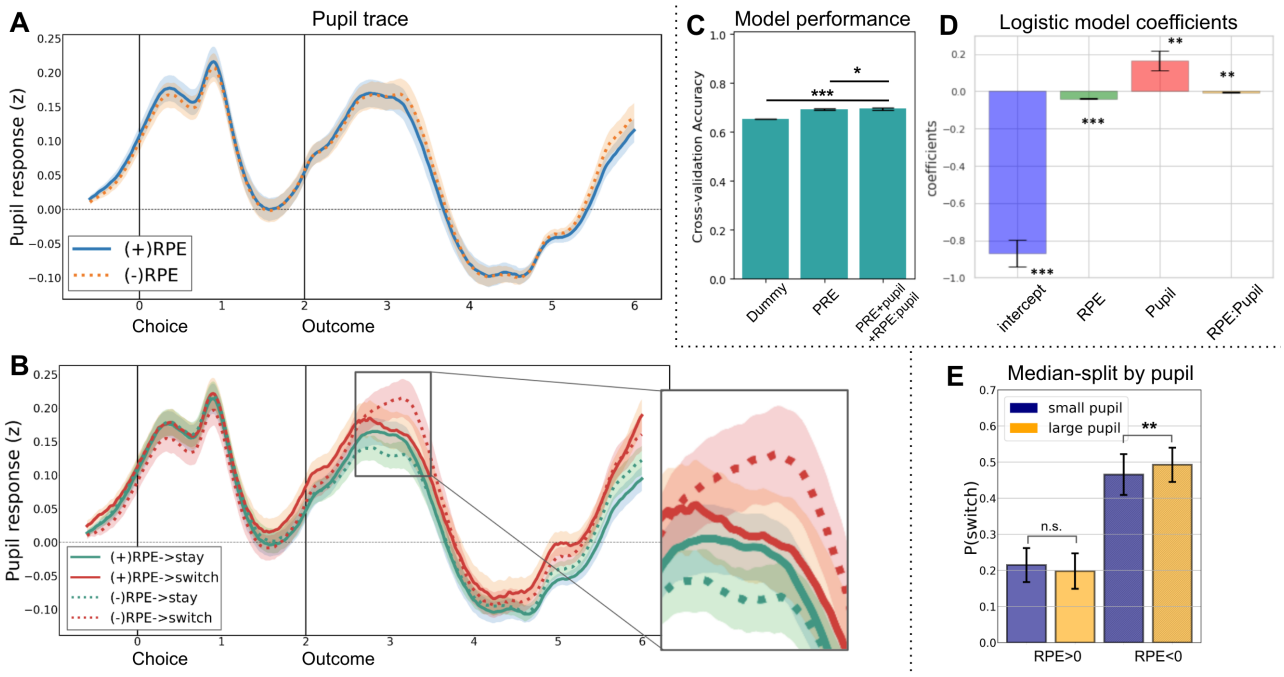

Figure 2: Interaction effect of RPE and choice on pupil response. (A-B) Z-scored baselined pupil response trace over the course of the outcome-viewing period, separated by (A) sign of RPE, and (B) both RPE sign and upcoming choice (paired-t test by subjects: $p_{1,3} = 0.01$, $p_{1,4} < 0.001$, $p_{2,4} < 0.001$, $p_{3,4} < 0.001$, where 1,2,3,4 correspond to the order of the legend). (C) Model fit (accuracy) across logistic regression models predicting choice on the next trials. (D) Logistic regression coefficients for RPE, post-outcome pupil response, and interaction term. (E) Average frequency of staying on the next trial given possible or negative RPE, trials median-split by pupil size. Frequency of staying given negative RPE is larger in the condition with larger pupil size (paired t-test, $p < 0.001$); frequency of staying given positive RPE is not significantly different for the two pupil conditions (paired t-test, p=0.211).

two other models: (1) a baseline "dummy" model that always predicts "stay" (since subjects show a strong repetition bias), and (2) a logistic regression model with only RPE as the predictor (Figure 2C). The model including pupil response and the interaction term has significantly higher average cross-validation accuracy than the model with only RPE (paired t-test stats=-3.225, p=0.03) and the dummy model (paired t-test stats=-17.71, $p < 0.001$). The regression coefficients (Figure 2D) are significant for both RPE, pupil response, and the interaction term. The results above suggest that post-outcome pupil response is predictive of upcoming choice, in a way that may be interactive with RPE.

To further quantify the dynamic RPE encoding in the pupil, we regress baseline-corrected pupil time courses against the trial-wise estimates of RPE magnitude ($|RPE|$) and EU difference between the two options ($\Delta EU = EU_{chosen} - EU_{unchosen}$), separately for positive and negative RPE, and for stay and switch trial conditions (Figure 3). The magnitude of pupil scaling differs depending on both the sign of RPE and the future choice action. When RPE is positive, pupil response significantly scales with the size of RPE only if the upcoming choice is a stay rather than a switch (Figure 3A, top figure). In contrast, when RPE is negative, pupil response shows a larger scaling with $|RPE|$ if the upcoming choice is switch rather than stay (Figure 3A, bottom figure). In other words, the predictability of pupil trace on the magnitude of RPE changes across the choice conditions. Thus, what the pupil response reflects is not simply the value or magnitude of RPE or a signal indicating switch versus stay; rather, it reflects the relationship between RPE and the upcoming choice. Larger encoding of the magnitude of RPE in the pupil size during the reward feedback (in both positive and negative RPE conditions) leads to more reward and error-driven choices (i.e. to stay when the reward increases and switch when the reward decreases). The significant scaling difference emerges shortly (0.5 to 1s) after the outcome onset, suggesting that future choice difference is reflected in the early feedback processing stage.

We perform another regression analysis on the pupil time course against the trial-wise difference in EU ($\Delta EU = EU_{chosen} - EU_{unchosen}$, Figure 3B). When RPE is negative, pupil size scales more positively with ΔEU if followed by a stay choice (Figure 3B, top). When RPE is positive, pupil size shows moderately but not significantly larger scaling with ΔEU before a switch choice (Figure 3b, bottom). As with $|RPE|$, the pupil response seems to encode not the value but

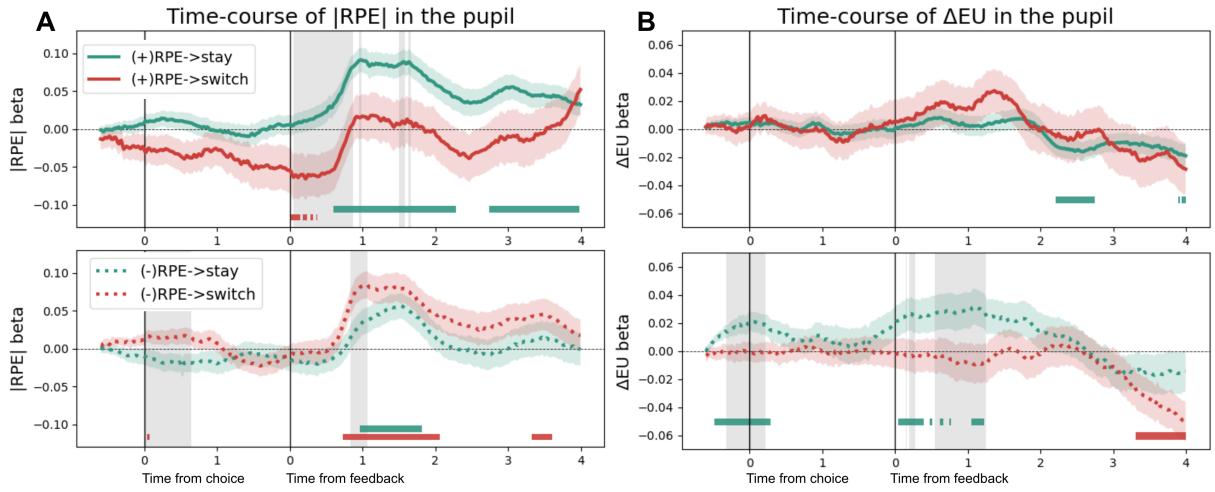

Figure 3: Time course of **A.** RPE and **B.** ΔEU ($EU_{chosen} - EU_{unchosen}$) scaling in the pupil, computed as the sample-by-sample multiple linear regression of baseline-corrected pupil dilation onto the two variables. Regression coefficients were computed separately for the four choice conditions. RPE and ΔEU were normalized for each subject before splitting into different conditions. Lower bars (green and red) indicate p-values ≤ 0.05 from the Wilcoxon signed-ranked test of the difference between each time course and zero. Shaded areas (gray) indicate p-values ≤ 0.05 from a two-sample paired t-test between the two time courses. Baseline: 100ms before the trial onset.

the relevance of ΔEU in forming the upcoming choice: larger ΔEU encoding in the pupil size leads to more uncertainty-driven and less reward-driven choice (e.g. staying on the more uncertain choice even after a worse-than-expected outcome) in the following trial.

RPE & EU-related pupil response reflects individual variability in performance and choice bias

To understand how variabilities in pupil encoding of RPE and EU across the four choice conditions might contribute to different overall decision behavior, as well as performance differences across subjects, we take the subject-wise regression coefficients for $|RPE|$ and ΔEU shown in Figure 3 for each condition, and correlate them with individual subjects' tendency to stay (Figure 4). After experiencing a negative RPE, subjects whose pupil responses were more predictive of $|RPE|$ before a (-)RPE⇒switch action, and less predictive of RPE before a (-)RPE⇒stay action, are generally more likely to stay and had better overall task performance (Figure 4A).

Furthermore, subjects with larger pupil encoding of negative RPE before a (-)RPE⇒switch action also have lower subjective volatility rates (Pearson's r = -0.31, p=0.023), a model-fitted subjective belief of rate of reward change (see Method) that is associated with self-reported lower anxiety scores in State-Trait Anxiety Inventory (STAI-Y) test (STAI-Y-A: Pearson's r=0.32, p=0.017; STAI-Y-B: Pearson's r=0.29, p=0.035).

After experiencing a positive RPE, larger pupil encoding of ΔEU before a (+)RPE⇒switch action is associated with a higher frequency of staying and better performance (Figure 4B). Together, we see that individual differences in choice tendency, partially associated with subjective beliefs of environmental volatility and anxiety levels, are reflected in variable pupil encoding of RPE and EU to guide upcoming choices. More conservative subjects show larger encoding of RPE to guide reward and error-driven switch actions ((-)RPE⇒switch) and larger encoding of *Delta*EU to guide uncertainty-driven switches ((+)RPE⇒switch).

Discussion

In this study, we explore the interplay between the pupil-linked arousal system and adaptive value-based decision-making amidst uncertain and fluctuating environments. Utilizing a novel two-armed bandit paradigm with intermittent shifts in reward contingencies, our findings reveal that pupil dynamics are not mere reflections of reward and uncertainty values based on prior experiences. Instead, they signify how these factors influence future choices when performing adaptive decision-making in a dynamic environment. Larger pupil encoding of reward prediction error (RPE) facilitates reward-driven switches in choice, while larger pupil encoding of estimation uncertainty (EU) facilitates uncertainty-driven switches in choice. Furthermore, individual differences in pupil's encoding of size of RPE and EU difference correlate with individual variability in choice tendency and task performance. More conservative subjects show larger encoding of RPE to guide reward and error-driven switch actions ((-)RPE⇒switch) and larger encoding of *Delta*EU to guide uncertainty-driven switches ((+)RPE⇒switch).

Our investigation extends the theoretical framework (Yu & Dayan, 2005) that proposes two forms of task-related uncertainty crucial for effective learning and decision-making in

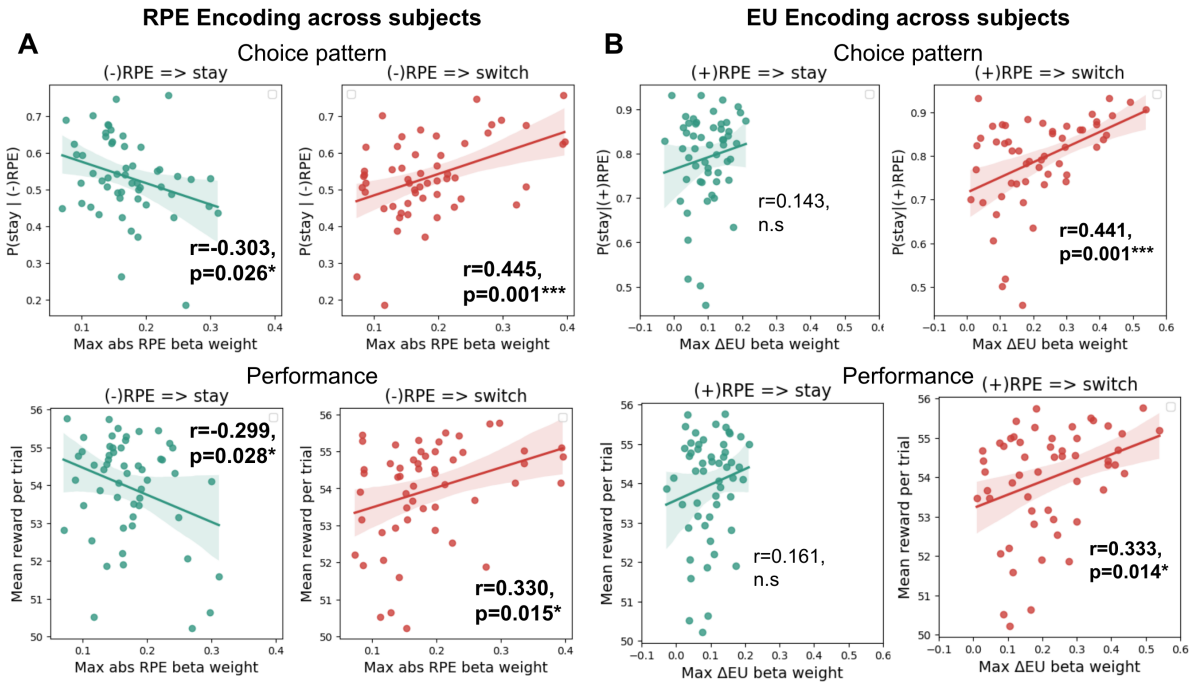

Figure 4: Individual choice bias (top) and task performance (bottom), correlated with per-subject RPE and ΔEU beta weights computed in Figure 3. **A.** Correlation with RPE weights for $(-)RPE \Rightarrow stay$ condition (green) and $(-)RPE \Rightarrow switch$ condition (red). **B.** Correlation with ΔEU weights for $(+)RPE \Rightarrow stay$ condition (green) and $(+)RPE \Rightarrow switch$ condition (red).

dynamic settings: expected uncertainty, stemming from observational noise and a known ignorance about the environment (as captured by estimation uncertainty in our study), and unexpected uncertainty, arising from unpredicted, gross changes in the environment (as reflected by reward prediction errors). Our results underscore how these uncertainties are encoded in the pupil's dynamics, influencing adaptive decision-making processes.

In line with the theory, a larger pupil encoding of RPE facilitate a shift towards reward-driven choices, mirroring the role of unexpected uncertainty in prompting individuals to reconsider and potentially revise their strategies and beliefs in light of new outcomes (Cohen et al., 2007). On the other hand, when the pupil predominantly encodes EU, which aligns with the concept of expected uncertainty, subsequent choices appear less influenced by recent reward outcomes, indicating a strategy to reduce known environmental uncertainties, enhancing long-term learning and reward gain. Given the known modulation of pupil size by norepinephrine (NE) and acetylcholine (ACh) systems (Joshi & Gold, 2020; Mathôt, 2018; Reimer et al., 2016; Gilzenrat et al., 2010; Aston-jones & Cohen, 2005), our findings also resonate with the proposed antagonistic interplay between NE and ACh in modulating adaptive cognitive processing (Cohen et al., 2007; Yu & Dayan, 2005). The distinct and variable pupil encodings of RPE and EU in our study can reflect the underlying neural mechanisms of NE and ACh in guiding adaptive decision-making.

Our findings can also be interpreted in the context of adaptive gain theory related to the pupil-linked LC-NE system (Aston-jones & Cohen, 2005). The theory suggests that the firing activities of LC-NE neurons modulate cortical circuit responsiveness, facilitating task engagement or disengagement and thus enabling adaptive responses to dynamic environments. Our findings suggest that the pupil-linked LC system may regulate engagement with options by modulating the informational value of rewards and uncertainties during decision-making. This modulation appears to direct upcoming choices to be more reward-focused or uncertainty-focused, depending on the contextual relevance of each. Additionally, individuals with a stronger propensity for repetition require a greater enhancement in information gain from rewards and uncertainty to override this bias and adapt their choices accordingly. This nuanced understanding of the LC-NE system's role offers a bridge between learning mechanisms and adaptive decision-making.

The reason behind the system's dynamic adjustment in prioritizing reward and estimation uncertainty remains an open question. A potential explanation, rooted in the efficient coding theory (Barlow, 1961), is that the pupil-linked LC-NE arousal system might be instrumental in the optimal distribution of cognitive resources during adaptive decision-making. This hypothesis aligns with the observation that decision-making leverages both reward and uncertainty to efficiently allocate cognitive effort, particularly in contexts where such allocation yields significant adaptive advantages. The ten-

dency to explore in the face of estimation uncertainty aligns with a strategic investment in gathering information for long-term gain, while the shift in choice following large negative reward prediction error after change-points underscores an adaptive response to unexpected environmental changes.

However, our study does not establish a causal link between pupil size variations and adaptive choice behavior. Future research should, therefore, employ methods such as manipulating pupil size via luminance adjustments to directly assess its impact on decision-making. Such investigations could further unravel the complex interplay between cognitive processes and decision-making under uncertainty, enriching our understanding of the brain's navigational strategies in complex decision landscapes.

Methods

Dataset

A total of sixty French adult individuals, free from any known psychiatric or neurological disorders, participated in this study in exchange for a monetary reward. Two participants were excluded due to incomplete participation in the experiment. Additionally, one participant was excluded following the incidental discovery of a brain anomaly during MRI scanning. A further three participants were excluded due to chance-level task performance (the average reward obtained fell below two standard deviations above chance level, which was determined by the average reward accrued across 1,000 iterations of making random choices on the same reward sequences). After these exclusions, the data from the remaining fifty-four subjects were analyzed. This group had an average age of 25.9 years (SD = 6.5, age range: 18-44), consisting of 28 women and 26 men.

Experimental Design

Novel two-armed bandit task Participants performed 8 blocks (2 hours long) of a novel non-stationary two-armed bandit decision-making task. Each block involves 96 trials of repeated choices among two options. The reward outcome for each option (ranges between 1 and 100) is sampled from a Gaussian distribution with a mean that switches among three values (30, 50, and 70) at random intervals of change points based on a true volatility rate of 1/24 (change happens every 24 trials on average, with a minimum of 10 trials between consecutive changes on the same option). The occurrences of change-points are independent for the two options. The standard deviation of the reward generation process for both options changes between periods of 10 (low noise) or 20 (high noise), and subjects are explicitly informed about the current noise level on every trial. Each game includes periods of 4 consecutive forced-choice trials to independently control for reward and uncertainty levels.

Pupil diameters are recorded from the subjects during the task using Tobii at 60Hz. The subjects are also asked to complete clinical questionnaires (BIG-5, STAI, Lot R) for anxiety and pessimism assessments.

Learning model: Bayesian Ideal Observer (IO)

We assume that subjects behave as an ideal Bayesian observer, tracking the reward value and uncertainty in their estimation. We denote the reward obtained from arm i ($i = 0$ denotes the left arm, $i = 1$ denotes the right arm) on trial t as $y_{i,t}$, which generated from a Gaussian distribution with mean $\mu_{i,t} \in \{30, 50, 70\}$ and standard deviation $\sigma_t \in \{10, 20\}$. With probability v (volatility), a change-point occurs on arm i and $\mu_{i,t+1}$ is redrawn from $\{30, 50, 70\}$ without replacement ($\mu_{i,t+1} \neq \mu_{i,t}$). Change-points happen independently for each arm. The goal of the subjects is to infer the trial-by-trial underlying reward mean μ_i of the two arms without knowing when the change-point occurs. By Bayesian probability theory, the posterior estimate of the reward mean $\mu_{i,t+1}$ of arm i on trial $t+1$, given all past observations $y_{1:t+1}$, the known standard deviations $\sigma_{1:t+1}$, and volatility v, is given by:

$$p(\mu_{i,t+1}|y_{1:t+1}, \sigma_{1:t+1}, v) \propto p(\mu_{i,t+1}, y_{1:t+1}, \sigma_{1:t+1}, v)$$
$$\propto p(y_{t+1}|\mu_{i,t+1}, \sigma_{1:t+1}) \sum_{\mu_{i,t}} p(\mu_{i,t}|y_{1:t}, \sigma_{1:t}, v) p(\mu_{i,t+1}|\mu_{i,t}, v)$$

Volatility v is fitted for each subject as a free parameter in the Bayesian model. Using the ideal observer model, we derive the trial-wise estimates of the following variables:

Expected Reward (ER) the expected value of the posterior estimate of the reward mean, marginalized over the posterior distributions of the three levels of reward mean: $ER_{i,t} = E[\mu_{i,t}]$

Reward Prediction Error (RPE) difference between the actual reward obtained and the expected reward: $RPE_{i,t} = R_{i,t} - ER_{i,t}$

Estimation Uncertainty (EU) the posterior probability that the most likely reward mean is not the actual generative reward mean:

$$EU_{i,t} = P(\mu_{i,t} \neq argmax_r P(\mu_{i,t} = \mu|y_{1:y})|y_{1:t})$$

Pupil Diameter Measurements

Pupil diameter was sampled at 60 Hz and recorded throughout the task using Tobii eye-tracker. Blinks detected by the tracker device's detection algorithm were removed using linear interpolation of values measured before and after each identified blink with a margin of 0.1. Blink-filtered diameter was low-pass filtered using a bi-directional 4th order Butterworth filter with a cutoff frequency of 30 Hz.

The pre-processed pupil measurements were then z-scored in each session for each participant. The average pupil size after outcome was computed for each trial by taking the mean of z-scored pupil measurements during the $[0.5, 1.5]$ second interval of the outcome-viewing period. The change of the pupil response after outcome is obtained by subtracting the baseline pupil average (100ms before the outcome onset) from the pupil measurements.

References

Aston-jones, G., & Cohen, J. D. (2005). An Integrative Theory of Locus Function : Adaptive Gain and Optimal Performance.
doi: 10.1146/annurev.neuro.28.061604.135709

Barlow, H. B. (1961). Possible Principles Underlying the Transformations of Sensory Messages. In *Sensory communication* (pp. 216–234). MIT Press. doi: 10.7551/mitpress/9780262518420.003.0013

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933–942. doi: 10.1098/rstb.2007.2098

Colizoli, O., de Gee, J. W., Urai, A. E., & Donner, T. H. (2018). Task-evoked pupil responses reflect internal belief states. *Scientific Reports*, *8*(1), 1–13. doi: 10.1038/s41598-018-31985-3

Fan, H., Burke, T., Sambrano, D. C., Dial, E., Phelps, E. A., & Gershman, S. J. (2023). Pupil Size Encodes Uncertainty during Exploration. *Journal of cognitive neuroscience*, *35*(9), 1508–1520. doi: 10.1162/jocn$_{a0}$2025

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 and Behavioral Neuroscience*, *10*(2), 252–269. doi: 10.3758/CABN.10.2.252

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration-exploitation tradeoff: Evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, *23*(7), 1587–1596. doi: 10.1162/jocn.2010.21548

Joshi, S., & Gold, J. I. (2020). Pupil size as a window on neural substrates of cognition. *Trends in Cognitive Sciences*, *24*(6), 466–480. doi: 10.1016/j.tics.2020.03.005

Kucewicz, M. T., Dolezal, J., Kremen, V., Berry, B. M., Miller, L. R., Magee, A. L., ... Worrell, G. A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific Reports*, *8*(1), 1–7. Retrieved from http://dx.doi.org/10.1038/s41598-018-23197-6 doi: 10.1038/s41598-018-23197-6

Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. , *1*(1), 1–23.

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*(7), 1040–1046. doi: 10.1038/nn.3130

Preuschoff, K., 't Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, *5*(SEP), 1–12. doi: 10.3389/fnins.2011.00115

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, *7*(May), 1–7. Retrieved from http://dx.doi.org/10.1038/ncomms13289 doi: 10.1038/ncomms13289

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*. doi: 10.1038/ncomms14637

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin and Review*, *25*(6), 2005–2015. doi: 10.3758/s13423-018-1432-y

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans Use Directed and Random Exploraiton to Solve the Explore-Exploit Dilemma. *Journal of experimental psychology. General*, *143*(6), 2074–2081. doi: 10.1037/a0038199

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692. doi: 10.1016/j.neuron.2005.04.026