Role of pupil-linked uncertainties and rewards in value-based decision making

Zoe He, Dalin Guo, and Angela Yu*
(wah016@ucsd.edu, dag082@ucsd.edu, ajyu@ucsd.edu)
Department of Cognitive Science, University of California San Diego, San Diego, CA, USA

Maëva L'Hôtellier, Alexander Paunov, and Florent Meyniel*
(maeva.l.hotellier@ens.psl.eu, alexander.paunov@cea.fr, florent.meyniel@cea.fr)
INSERM-CEA Cognitive Neuroimaging Unit (UNICOG), NeuroSpin Center, CEA Paris-Saclay, Gif-sur-Yvette, France Université de Paris, Paris, France

Abstract

While the role of uncertainty in learning and inference has been well studied, its additional contribution to exploratory decision making is less well understood. Here, we investigate through behavioral and pupillometric data how uncertainty and reward jointly govern exploration in decision making in a novel two-armed bandit paradigm, in which there are occasional changes in reward contingencies (change points). In terms of expected uncertainty (EU), which is thought to reflect observation noise and inferential uncertainty, we find subjects accord EU an exploratory value that is amplified when expected reward (ER) is larger. In terms of unexpected uncertainty (UU), which is associated with detecting change points, we find that subjects readily switch choices after change points, and the magnitude and timing of the transient pupil response reflect when the behavioral switch happens. These findings reveal pupillometry measures as a valuable tool for revealing uncertainty- and reward-related factors driving learning and exploration.

Keywords: Uncertainty; Reward; Pupillometry; Value-Based Decision Making; Exploration

Introduction

Humans and animals often need to make value-based choices in an imperfectly known and unstable environment, where the decision makers face an "exploration-exploitation" dilemma: they have to choose between the currently best option and explore a lesser known alternative. In order to successfully make inferences about the environment and choose among options, decision makers need to handle different sources of uncertainty: e.g. uncertainty due to external noise and our ignorance of the environment (expected uncertainty, or EU), and uncertainty that arises from unexpected, gross changes in the environment (unexpected uncertainty, or UU) (Yu & Dayan, 2003, 2005).

The role of uncertainty in learning has been extensively studied: EU and UU are believed to be signaled separately by cortical Acetylcholine (ACh) and Norepinephrine (NE) activities to regulate the influence of incoming information on existing beliefs (Yu & Dayan, 2003, 2005), and EU and UU are thought to be distinctly represented in baseline pupil size and transient pupil dilation, respectively (Nassar et al., 2012; Reimer et al., 2016). However, the role of uncertainty in the decision policy remains less well understood (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006). Here, we use behavioral measures and pupil metrics in a novel two-armed bandit task containing change points to examine how different forms of uncertainty and reward govern exploration in decision making.

Method

Behavioral experiment

Thirty-six subjects performed 8 sessions of 96 trials in a novel two-armed bandit decision-making task that involved repeated choices among 2 options. Each game includes periods of 4 consecutive forced-choice trials in order to independently control for reward and uncertainty levels.

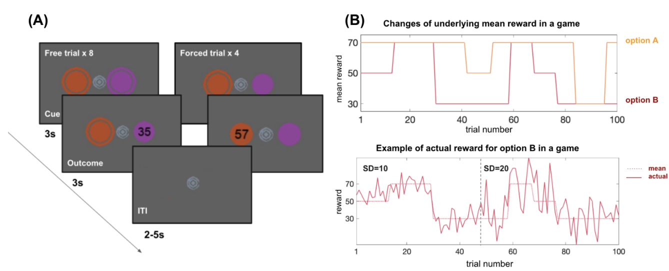

Figure 1: Experimental settings. (A) Game display. (B) Reward generation process visualized.

The reward outcome for each option 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 of 1/24. The occurrence 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.

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

Results

UU mediated pupil dilation modulates post-change-point switching / exploratory behavior. Behaviorally, subjects switch 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 (Fig. 2A). Interestingly, the magnitude of the transient pupil response (pupil change, thought to reflect UU) also peaks and then decays shortly after the high-to-low change point, with the peak occurring earlier in the low-noise condition (Fig. 2B). This behavior-pupil correlation suggests that change-point evoked pupil change response might reflect either a role for learning or exploratory behavior linked to the UU-NE system.

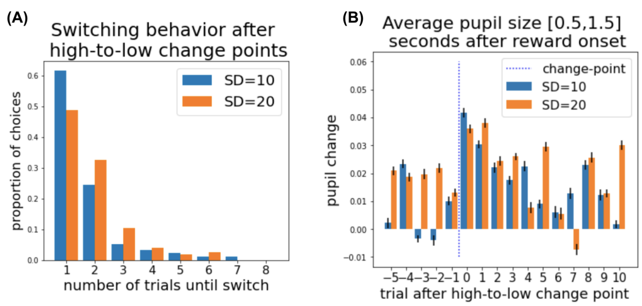

Figure 2: (A) Distribution of the number of trials until the first switching behavior after a high-to-low change point. (B) Average pupil change on trials around high-to-low change points (baseline: [-0.2, 0]s relative to stimulus onset).

UU mediated pupil dilation separately influences both the learning and exploration process. The magnitude of pupil dilation is larger for low-to-high change points on forced trials (Fig. 3A), but is larger for high-to-low change points on free trials (Fig. 3B). This forced-free difference may imply distinct roles of pupil dilation (UU) in learning and exploration policy: on force trials, after observations of changes from low to high reward, UU drives an increase in learning rate to update belief about a potential good choice; on free trials, after experiencing a drop in mean reward, UU may play a role in facilitating exploratory behavior to look for the new and better solution.

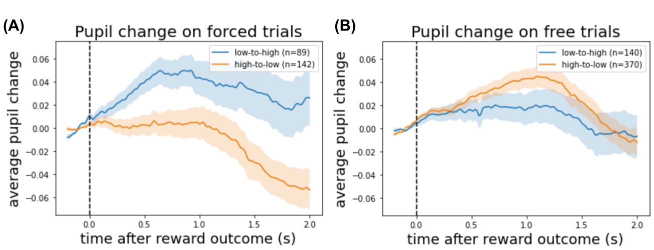

Figure 3: Pupil change to low-to-high (blue) vs. high-to-low (orange) change points on (A) forced trials. (B) free trials.

Interaction between expected reward and uncertainty. After a block of forced trials, subjects' first free choice appears to buy both the observed reward outcome and the unknownness/uncertainty of the options. While the option with lower mean reward outcome is generally less preferred, the tendency of choosing the low reward option increases when the option has higher uncertainty (observed fewer times). On the other hand, the tendency of choosing the option with higher mean reward outcome becomes even larger when the uncertainty level is high, suggesting a bigger role of uncertainty. Hence, EU likely interacts with expected reward multiplicatively rather than additively to determine choice.

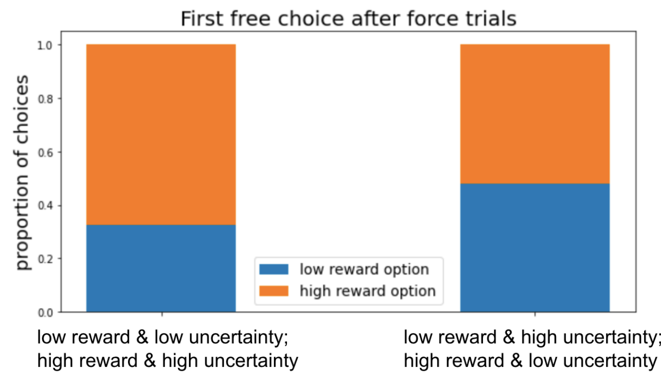

Figure 4: Figure 4: Fraction of choices allocated to the low reward options (blue) and the high reward options (orange).

Discussion

In short, pupil change, which reflects unexpected uncertainty, may signal learning and exploratory choices distinctly after a change point. The role of uncertainties in exploration interacts with the expected rewards of the options. These findings have important implications on the potential normative framework for decision policies involving uncertainties: the contribution of uncertainties may not be additive, but possibly multiplicative, interacting with expected reward.

Acknowledgments

This work is funded by a CRCNS grant to AY (NIH- R01 DA050373) and FM (ANR 19-NEUC-0002-01).

References

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879. doi: 10.1038/nature04766

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

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

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

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