Gains and Losses Modulate Novelty-Seeking During Explore-Exploit Decisions

Kathryn M. Rothenhoefer (rothenhk@ohsu.edu)
Division of Neuroscience, Oregon National Primate Research Center, Oregon Health & Science University, Portland, OR, USA

McKenna D. Stocker (stockerm@ohsu.edu)
Division of Neuroscience, Oregon National Primate Research Center, Oregon Health & Science University, Portland, OR, USA

Vincent D. Costa (costav@ohsu.edu)
Division of Neuroscience, Oregon National Primate Research Center, Oregon Health & Science University, Portland, OR, USA

Abstract:

Explore-exploit decision making requires balancing exploiting known options and exploring novel options in order to maximize long term value. Understanding valence dependencies in explore-exploit tradeoffs remains a challenge, as it is difficult to equate appetitive and aversive primary reinforcers. We solve this problem by using virtual tokens as secondary reinforcers to be exchanged for juice rewards. Rhesus macaques start with a token endowment, and then choose between novel and familiar cues during a multi-arm bandit task where cues are associated with either a gain or loss of tokens. Monkeys efficiently learn to approach cues associated with gains and avoid cues associated with losses, however, directed exploration of novel choice options was dependent on the reward horizon and valence context. These results imply that primates value information over rewards, especially when attempting to resolve uncertainty in aversive decision contexts.

Keywords: novelty; explore-exploit decision making; reward; valence

Introduction

Successful decision making requires that humans and animals balance exploiting known options, and exploring novel options in order to maximize long term value (Daw et al., 2006; Sutton & Barto, 2018). This tradeoff is known as the explore-exploit dilemma. Explore-exploit tradeoffs have been studied extensively in the context of maximizing gains (Costa et al., 2019; Wilson et al., 2014), or minimizing punishments (Krueger et al., 2017). Humans and nonhuman primates are novelty-seeking in the context of maximizing gains (Costa et al., 2019; Hogeveen et al., 2022), but how primates resolve the explore-exploit dilemma when exploration potentially leads to loss or punishment remains poorly understood. While it is adaptive to orient to novel stimuli in order to assess whether they are threatening (Bradley, 2009), the primacy of aversive reinforcers appears to influence novelty driven exploration in aversive contexts (Lejarraga & Hertwig, 2017) (Bublatzky et al., 2017).

Computational solutions to managing explore-exploit trade-offs have not explicitly considered valence dependencies in decision making. This reflects a conceptual bias that exploration is an intrinsically appetitive act and that information is always desirable. One reason for bias are the difficulties in finding equivalent appetitive and aversive outcomes that elicit comparable approach and avoid behaviors. This is especially true studying decision making in nonhuman primates. In order to create a task where macaques can gain and lose in the same currency, we utilized tokens as a secondary reinforcer, to be exchanged for juice rewards. This establishes an equivalent currency across valence domains, allowing us to investigate novelty-driven exploration in scenarios where choices are associated with different size gains or losses, and a token endowment that needs to reach a threshold before juice exchange is guaranteed.

Results

Two rhesus macaques were trained to complete a three-arm bandit task to acquire tokens that were cashed out for juice reward (Fig. 1a). Animals are given an 8 token endowment to start and after every cash out. Following a choice of one of three cues, if the chosen cue resulted in a gain, tokens were added to the screen, and if the cue resulted in a loss, tokens were removed. If the animals accrued 4 to 11 tokens, they had a linearly increasing chance of cashing out each token for a 0.1 ml drop of juice. If they had 12 or more tokens they were guaranteed to get cashed out. Each block consisted of approximately 10-30 trials, in which the monkey had to choose among three images that are drawn from a set of four for the block (Fig. 1b). Each cue was assigned a mean value in terms of tokens gained or lost. The

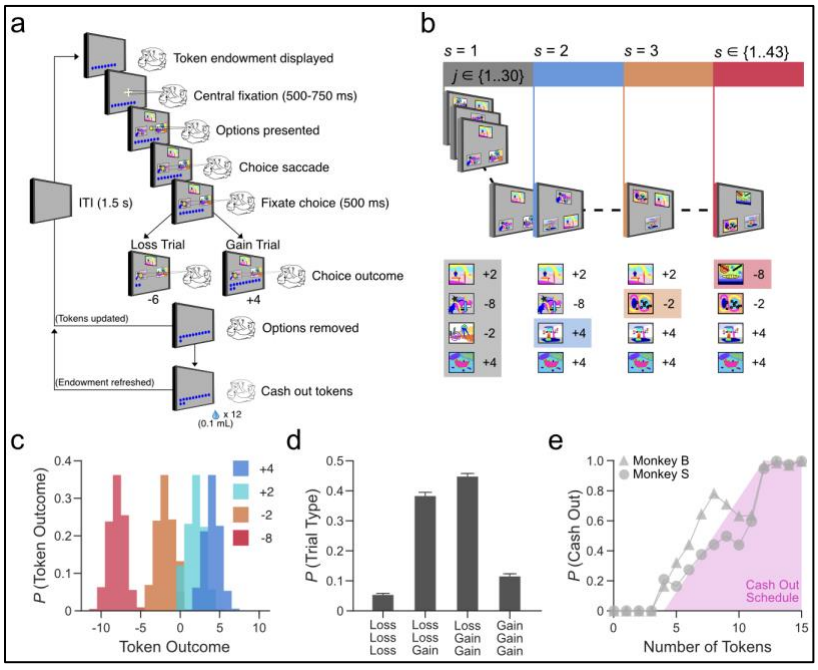

Figure 1: Task. (a) Trial structure. **(b)** Block structure within a session. **(c)** Distribution of token outcomes for each cue average. **(d)** Probability of each combination of gain and loss cues. **(e)** Experienced probability of cashing out their tokens by the number of tokens at the end of a trial for monkeys B and S (grey), and the programmed cash out schedule (pink).

number of tokens gained or lost on each trial following selection of a cue was drawn from a normal distribution surrounding its mean value. The mean token distribution values were -8, -2, +2, and +4 (Fig. 1c). The monkeys were more likely to experience trials where there it was possible to gain rather than lose tokens (Fig. 1d; Trial Type: $F(1,320) = 9.57$, $p < 0.01$). Figure 1e shows the intended cash out schedule (pink shaded region), and the actual cash out probabilities for monkeys B and S.

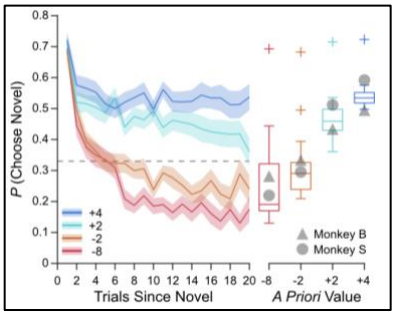

Figure 2: Choice Behavior. Probability of choosing the novel cue as a function of trials since the novel cue was introduced, separated by novel cue value.

In order to determine the effect of value on novelty-seeking, we averaged the probability of choosing the novel option as a function of the trials since the novel cue was introduced, separated by the value of the novel option (Fig. 2). There was a significant effect of valence on the probability of choosing the novel option ($F(1,105) = 692.96$, $p < 0.001$). Furthermore, there was a significant difference in choosing between the loss cues ($F(1,105) = 47.2$, $p < 0.001$) and the gain cues ($F(1,105) = 32.76$, $p < 0.001$). This result indicated that the animals were able to learn and differentiate between cues that predicted varying amounts of token gains and losses, and make rational decisions.

Previous work has shown that the value of the more familiar option, and proximity to receiving a cash out, plays an important role in whether the decision maker will seek more information about an unknown option (in our case, the novel option) (Wilson et al., 2014). Figure 3a illustrates that the monkeys are less likely to choose the novel option when it is introduced as the value of the best alternate cue increases – even before they have sampled the novel option to determine its value (Valence: $F(1,112) = 25.74$, $p < 0.001$). Previous work has shown that as human decision makers get closer to the cashout window, they become less likely to select the novel option, and more likely to select the best alternative option (Wilson et al., 2014). In other words, when they have a short horizon (closer to cash out) they are more novelty-averse than when they have a long horizon (further to cash out). To investigate this effect, we utilized the model from the aforementioned study that investigated the effect of reward horizon and information on decision making (Wilson et al., 2014). We estimated an information bonus for choosing the more informative option (novel cue) for each monkey 1,000 times for all possible token amounts leading up to 100% probability of cashout. We did the same procedure but with data that had been shuffled so that the tokens were random for each model fitting. We found that the information bonus estimates from the real data were significantly different from the shuffled data (p <0.01, 95% CI [-0.23, -0.14]). These results indicated that as the animals get closer to the cash out window, they value the information gained from choosing the novel cue less (Fig. 3b). Figure 3c replicates information bonus parameter differences seen in human decision makers (Wilson et al., 2014). These results are the first to replicate this human decision-making behavior in rhesus monkeys.

Conclusion

To summarize, this study provides the blueprint for using token economies to study the explore-exploit dilemma across gain and loss contexts in nonhuman primates (NHPs). This work leaves a clear path to future neurophysiological studies to determine the neural underpinnings that facilitate our ability to perform explore-exploit decision making across gain and loss contexts.

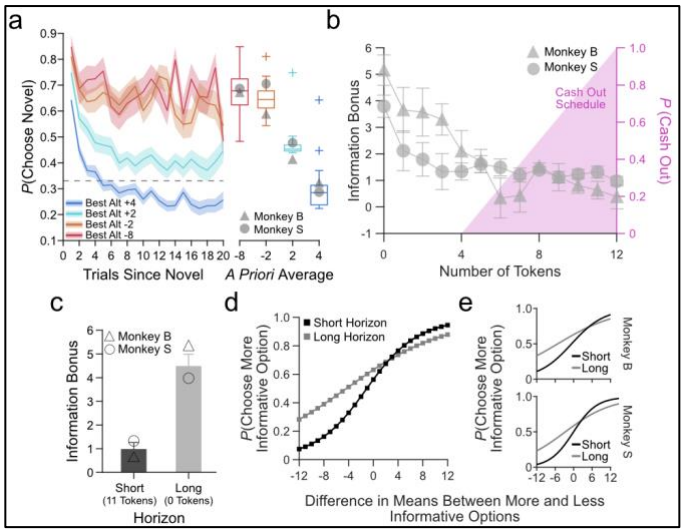

Figure 3: Effect of Cash Out Horizon on Behavior. (a) Probability of choosing the novel cue as a function of trials since the novel was introduced, separated by best alternate cue value. **(b)** Information bonus parameter as a function of number of tokens (grey). Pink shaded region illustrates the cash out probability as a function of number of tokens. **(c)** Information bonus parameters for a short horizon (11 accrued tokens) and a long horizon (0 tokens). **(d)** Model estimates of the probability of choosing the more informative option (the novel cue) over the less informative option (best alternate cue) as a function of the value of the novel cue minus the value of the best alternate cue, for both short and long horizons. **(e)** Same as **f**, but separated for each monkey.

Acknowledgments

This work was funded by R01MH125824, awarded to VDC.

References

Bradley, M. M. (2009). Natural selective attention: orienting and emotion. *Psychophysiology*, *46*(1), 1-11.

Bublatzky, F., Alpers, G. W., & Pittig, A. (2017). From avoidance to approach: The influence of threat-of-shock on reward-based decision making. *Behav Res Ther*, *96*, 47-56.

Costa, V. D., Mitz, A. R., & Averbeck, B. B. (2019). Subcortical Substrates of Explore-Exploit Decisions in Primates. *Neuron*, *103*(3), 533-545 e535.

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.

Hogeveen, J., Mullins, T. S., Romero, J. D., Eversole, E., Rogge-Obando, K., Mayer, A. R., & Costa, V. D. (2022). The neurocomputational bases of explore-exploit decision-making. *Neuron*, *110*(11), 1869-1879 e1865.

Krueger, P. M., Wilson, R. C., & Cohen, J. D. (2017). Strategies for exploration in the domain of losses. *Judgment and Decision Making*, *12*(2), 104-117.

Lejarraga, T., & Hertwig, R. (2017). How the threat of losses makes people explore more than the promise of gains. *Psychonomic bulletin & review*, *24*, 708-720.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

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. *J Exp Psychol Gen*, *143*(6), 2074-2081.