Individual differences in strategic exploration may reflect rational consideration of learning

Kate Nussenbaum (katenuss@princeton.edu) & Nathaniel Daw (ndaw@princeton.edu)
Princeton Neuroscience Institute
Princeton, NJ, 08544, USA

Abstract

Previous research has found that people explore strategically, increasingly forgoing immediate rewards to gain information when information has more value. While evidence for strategic exploration is robust *on average*, there is substantial heterogeneity in the extent to which individuals consider information value when deciding whether to explore or exploit. Here, we extended an existing task to examine individual differences in how the value of information influences exploration. In a large sample of adults (N = 163), we demonstrate that sensitivity to the value of information may reflect the extent to which information is *used* to guide future choices.

Keywords: exploration; value-guided learning

Introduction

People often face decisions between 'exploiting' known options to maximize immediate reward gain versus 'exploring' more uncertain options to gain information. Extensive prior research investigating these decisions has found that people explore *strategically*, such that they increase their tendency to explore more uncertain options when gaining information can improve future choices (Wilson, Geana, White, Ludvig, & Cohen, 2014; Rich & Gureckis, 2018). Prior work has measured strategic exploration by manipulating two distinct task properties: the *horizon* over which people will make future decisions (Wilson et al., 2014) and the contingency between choice and information (Rich & Gureckis, 2018). If people behave strategically, shortening decision horizons and removing the choice-information contingency should both lead to reduced exploration in favor of immediate reward maximization.

It is unclear, however, whether these two manipulations lead to convergent effects, particularly in developmental and clinical populations that exhibit substantial (and potentially informative) heterogeneity in their exploratory behaviors (Somerville et al., 2017; Harms et al., 2024; Zhuang, Niebaum, & Munakata, 2023; Smith et al., 2022). Importantly, normative computations of 'information value' assume that information will be learned and used optimally to guide future choices. Any suboptimalities in learning, if the chooser accounts for them, should reduce the effective value of information and produce what appear to be 'failures' to fully to modulate exploration based on models of ideal responses to these task manipulations.

Our goals in this study were twofold. First, we aimed to adapt and extend an existing strategic exploration task (Wilson et al., 2014) so that it could be used in future work with developmental and clinical populations. We shortened the task, framed it within a child-friendly narrative, and removed all explicit numbers to mitigate confounds induced by individual differences in mathematical abilities. Second, we aimed to examine how individual differences in strategic exploration related to individual differences in value-guided learning. By adding a counterfactual (cf) feedback manipulation, we could test the extent to which the learning assumptions implicit in theories of optimal information-seeking aligned with adults' behavior.

Methods

Young adult participants (N = 163, ages 18 - 30 years, recruited from Prolific) completed an adapted version of the 'horizons' task (Wilson et al., 2014) online.

Participants completed four blocks of 40 'games' during which they had to select between two trees to pick the largest apples. After selecting a tree, participants received an apple (depicted as a red circle), that remained on the screen for the duration of the game (Fig. 1). Participants were told that some trees tended to produce larger apples than other trees, and that they should try to select the trees with the largest apples.

In each task block, participants experienced both 'short-horizon' and 'long-horizon' games, which were interleaved in a random order. Each game began with four forced-choice trials, in which participants had to select (via key press) the tree on which a sloth appeared (Fig. 1). The forced-choice trials were distributed such that participants selected three apples from one of the trees and one apple from the other tree, in a randomized order. This manipulation imposed a difference in information between the two trees. Following the forced-choice trials, either one or four monkeys appeared in the center of the screen, indicating the number of free-choice trials that participants could make – one in short-horizon games, and four in long-horizon games. In long-horizon games, after each free choice, a monkey disappeared from the screen, such that the number of remaining monkeys on the screen indicated to participants their remaining number of free choices.

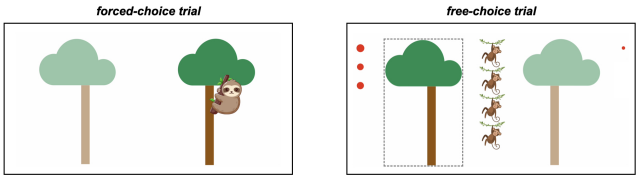

Figure 1: Participants completed 160 games in which they selected between two trees to pick the largest apples. Games comprised four forced-choice trials and either one or four free-choice trials.

Participants completed two 'baseline' blocks in which they only saw apples from selected trees, and two 'cf' task blocks in which they saw apples they *would have* received had they selected the alternative tree on free-choice trials. The motivation for these manipulations is that directed exploration (choice of the more informative option on the first trial) is predicted to increase in long-horizon games (because information gained can improve later choices) but this effect should disappear with cf feedback (where information does not depend on choice).

Results

Exploratory decision-making

We first examined participants' tendency to select the more informative option (i.e., the option with only 1 versus 3 apples picked from the forced-choice trials) on the first free-choice trial of every game, via a mixed-effects logistic regression with the experienced value difference between the trees, horizon, and block condition as interacting predictors. We hypothesized that in addition to making value-driven choices, participants would choose the more informative option more often in long vs. short-horizon games within baseline but not cf blocks.

In line with our predictions, we observed main effects of value, horizon, and condition in the directions we expected ($ps < .001$; Fig. 2). However, in contrast to our initial hypothesis, we did not observe a significant horizon x condition interaction effect, $\beta = .03, SE = .02, z = 1.8, p = .072$. Even when we examined cf trials on their own, we continued to observe a significant effect of horizon on information-seeking, $\beta = .07, SE = .02, z = 3.02, p = .003$. These data suggest that even when participants' choices had no influence on the information they received, they were still more likely to select the more uncertain option in long-horizon games.

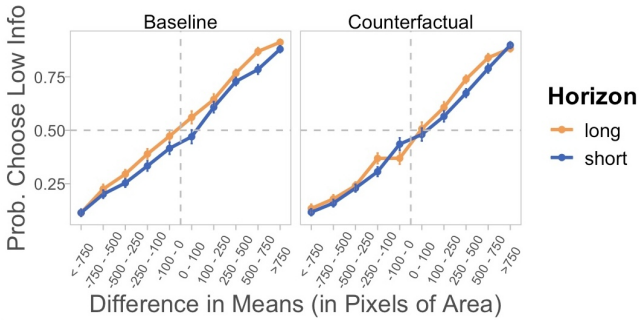

Figure 2: Participants more often chose the more informative option in long-horizon games and in baseline blocks.

Why did participants continue to demonstrate an effect of horizon on choice, even in the cf condition? One possibility is that participants do not learn as effectively from cf information (Li & Daw, 2011). Despite being presented with full feedback on every trial, they may still change their decisions more based on experienced outcomes, leading to greater effective benefits of 'information-seeking' in long-horizon games.

Value-guided learning

To test whether this was the case, we examined participants' subsequent free choices on long-horizon games in cf feedback blocks. We ran a logistic mixed-effects model examining the option participants selected as a function of experienced and cf outcomes, controlling for the overall mean of the outcomes revealed through forced-choice trials, the trial's 'more informative' option, and the previous' trial's choice. Participants' choices were influenced by all outcomes, though they

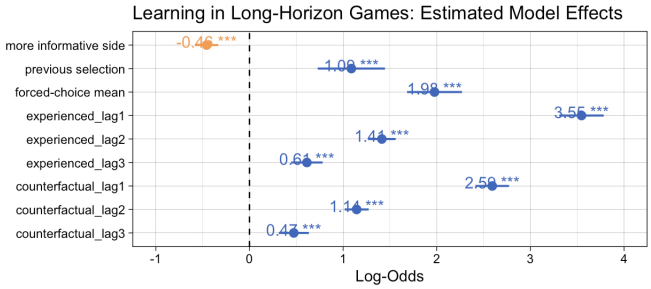

Figure 3: Participants learned more from experienced vs. cf outcomes.

weighted experienced outcomes more heavily (Fig. 3), indicating that persistent sensitivity to horizon in the cf feedback condition may reflect imperfect cf updating.

Relations between exploration and learning

We hypothesized that participants who learned the most from cf feedback would also show the greatest differences in information-seeking between baseline and cf feedback blocks. For each participant, we derived a 'cf learning ratio' by summing the cf feedback fixed effect from our learning model with their cf feedback random slopes, and dividing by the sum of their experienced feedback effects. On average, participants' cf learning ratio was below 1 (mean = .77 (SD = .16)).

We then ran a linear regression examining how these ratios related to individual differences in the effect of block condition (derived from our choice regression model) on information-seeking in long-horizon games. We found that participants who demonstrated the greatest differences in information-seeking between the baseline and cf blocks also demonstrated the best learning from cf feedback, $b = .06, SE = .02, t = 3.96, p < .001$.

Discussion

Our results demonstrate that our child-friendly task successfully reproduces signatures of strategic exploration in adult participants. However, in contrast to the predictions of normative models, the inclusion of cf feedback did not fully attenuate horizon effects on information-seeking.

There are additional possible reasons why participants may have shown information-seeking behavior even when information was not contingent on choice. Participants may have been more risk-seeking when they knew they had additional opportunities to make choices. Alternatively, rather than computing the value of information when faced with explore/exploit decisions, participants may have relied on simpler heuristic cues – like the game's 'horizon' – to determine how much to explore.

Here, we demonstrate preliminary evidence for a third possibility – participants' information-seeking may take into account their own subsequent learning biases. Thus, 'failures' to explore strategically may emerge from adaptive consideration of one's own future use of information.

Acknowledgments

We gratefully acknowledge our funders: the CV Starr Foundation (Fellowship to K.N.) and the NIMH (grant MH135587 to N.D., part of the CRNCS program).

References

Harms, M. B., Xu, Y., Green, C. S., Woodard, K., Wilson, R., & Pollak, S. D. (2024). The structure and development of explore-exploit decision making. *Cognitive Psychology*, *150*, 101650. doi: 10.1016/j.cogpsych.2024.101650

Li, J., & Daw, N. D. (2011). Signals in human striatum are appropriate for policy update rather than value prediction. *Journal of Neuroscience*, *31*(14), 5504–5511. doi: 10.1523/JNEUROSCI.6316-10.2011

Rich, A. S., & Gureckis, T. M. (2018). Exploratory Choice Reflects the Future Value of Information. *Decision*, *5*(3), 177–192. doi: 10.1037/dec0000074

Smith, R., Taylor, S., Wilson, R. C., Chuning, A. E., Persich, M. R., Wang, S., & Killgore, W. D. S. (2022). Lower Levels of Directed Exploration and Reflective Thinking Are Associated With Greater Anxiety and Depression. *Frontiers in Psychiatry*, *12*, 782136. doi: 10.3389/fpsyt.2021.782136

Somerville, L. H., Sasse, S. F., Garrad, M. C., Drysdale, A. T., Akar, N. A., Insel, C., & Wilson, R. C. (2017). Charting the expansion of strategic exploratory behavior during adolescence. *Journal of Experimental Psychology: General*, *146*(2), 155. doi: 10.1037/xge0000250

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*(6), 2074–2081. doi: 10.1037/a0038199

Zhuang, W., Niebaum, J., & Munakata, Y. (2023). Changes in Adaptation to Time Horizons Across Development. *Developmental Psychology*, *59*(8), 1532–1542. doi: 10.1037/dev0001529