

Present bias in exploratory choice

Alexander S. Rich (asr443@nyu.edu)

Todd M. Gureckis (todd.gureckis@nyu.edu)

New York University, Department of Psychology, 6 Washington Place, New York, NY 10003 USA

Abstract

Keywords time preference, present bias, exploratory choice

The problem of under-exploration seems remarkably common. One of the most severe expressions of under-exploration is learned helplessness, in which an organism experiences the absence of control over the environment, learns that the environment is uncontrollable, and thus ceases to take actions that might allow it to regain its sense of control. Learned helplessness has been proposed to underlie most notably depression (Abramson et al., 1978, 1989) but also problems ranging from difficulties in school CITE to poverty CITE.

While the cognitive appraisal of experienced events affects the development of learned helplessness (Abramson et al., 1978), patterns of exploration clearly play a role as well (Huys & Dayan, 2009; Teodorescu & Erev, 2014a). In the case of depression, interventions aimed at increasing the exploration of activities this might produce a sense of pleasure and mastery have been found to be as effective as those with a more cognitive orientation (Jacobson et al., 1996).

Under-exploration also seems to occur in the development of complex skills, such as flying a plane or playing a sport CITE. In these settings, an “emphasis change” training method in which the focus of practice is repeatedly changed can lead to greater performance gains than unguided practice or more complex training methods. Emphasis change works because it encourages people to continually explore the performance space to find more effective cognitive and motor routines CITE. Without these interventions, people often enter a “local maximum” in which exploration decreases and performance plateaus CITE.

There are many other areas in which under-exploration is less clearly established, but is suspected to play a role in maladaptive behavior. Insufficient exploratory interaction with outgroups may be one cause of stereotypes and prejudice (Denrell, 2005), and interventions that increase intergroup contact reduce stereotypes (Shook & Fazio, 2008). The crowding out of exploration by exploitation is a concern in organizational behavior as well (March, 1991; Levinthal & March, 1993), prompting research into organizational structures that may preserve exploration (Fang et al., 2010).

Under- and over-exploration in the lab

Given the seeming pervasiveness of under-exploration in real-world decision making, one might expect a clear pattern of under-exploration in lab experiments. Instead, however, lab results are mixed, with people sometimes under-exploring, sometimes over-exploring, and sometimes exploring close

to an optimal amount. To take two illustrative examples, Zwick et al. (2003) found that in a sequential search task people under-searched when there were no information costs but over-searched when there were information costs, and Teodorescu & Erev (2014b) found that people explored unknown alternatives too often or not often enough depending on whether rare outcomes were positive or negative. Similar results have been obtained within and across a variety of other studies and paradigms (Tversky & Edwards, 1966; Bussemeyer & Rapoport, 1988; Hertwig et al., 2004; Navarro et al., 2016; Juni et al., 2016; Sang et al., 2011).

These experimental studies raise the question of why under-exploration appears more widespread than over-exploration in the field, but not in the lab. One possibility is that both forms of deviation from optimality are in fact prevalent, though perhaps in different settings, and that the seemingly general bias toward under-exploration is illusory. However, in the current paper we propose instead that lab tests of exploratory choice have been missing a key component of real-world exploratory choice, namely, the distribution of choices and outcomes over time, and that this spreading of exploratory choices over time may account for the tendency to under-explore.

Exploration and myopia

Under-exploration is often proposed to be the result of myopia: the underweighting of temporally distant rewards relative to close ones (March, 1991; Levinthal & March, 1993). While choosing actions that are expected to immediately maximize reward is inherently exploitative, exploration often requires the decision-maker to suffer short-term costs in search of better future options.

There are two distinct possible causes of myopia (Bartels & Urminsky, 2015). The first is that the decision-maker is unaware of or fails to consider the future. This form of myopia is likely one cause of under-exploration, particularly in complex environments. However, researchers in exploratory choice have begun to address this form of myopia, and have found that people change their patterns of exploration based on their expectations about future choices, showing that, at least to some degree they do consider the future (Meyer & Shi, 1995; Wilson et al., 2014; Rich & Gureckis, 2017).

The second form of myopia is that the decision-maker is aware of the future but simply doesn't care as much about it. A wealth of studies have shown that this form of myopia is an important factor in human decision making, and that people are strongly biased towards present. For example, people will often prefer a larger, later monetary reward to a smaller, sooner reward when both rewards are in the fu-

ture. But when the time until both rewards is reduced so that the sooner reward is immediate, the sooner reward seems to loom larger and people may switch their preference (Kirby & Herrnstein, 1995). With non-monetary rewards, such as the cessation of an annoying noise (Solnick et al., 1980), watching a video when bored (Navarick, 1998), or drinking soda when thirsty (Brown et al., 2009), a bias towards immediate rewards has been observed on the scale of minutes or even seconds.

Present bias has the clear potential to cause people to explore too little, by increasing the perceived immediate gains of exploiting relative to the delayed gains of exploring. However, to date there has been no experimental research examining how people's bias towards the present affects exploratory choice. In part, this is a product of limits in experiment design. In most exploratory choice experiments, each choice and outcome takes fewer than 10 seconds to complete, making it unlikely for a large bias to develop between the current trial and future ones. More importantly, regardless of the length of each trial, rewards are generally not consumed on a trial-by-trial basis. For example, a participant might receive "points" on each trial, which are converted to money when the experiment ends. In this sort of paradigm, there is no reason to expect a bias towards the present choice's rewards over future rewards, as all rewards are in fact delivered at the same time.

In the present study, we describe our ongoing efforts to address the lack of research on present bias and exploratory choice. We develop a novel exploratory choice task using real consumption rewards and present two experiments. In the first, we vary the expected payoff of exploration across conditions and test whether this affects participant choice, to ensure that participants understand the task and respond as predicted to well-studied variables. In the second, we introduce a delayed-outcome condition, an intervention only expected to increase exploration if people are present biased. While we find no effect of delay on people's choices, we discuss plans for future experiments as well as the difficulties of studying choice with consumable rewards.

Experiment 1

Method

Participants. Thirty-one participants completed the experiment, which was conducted over Amazon Mechanical Turk using the psiTurk framework (Gureckis et al., 2015). Participants were paid \$5.00 for their participation, with a performance-based bonus of up to \$3.00. All participants received the full \$3.00 bonus.

Design and procedure. Participants were informed that their job was to perform a monotonous slider task that would be split into 30-second "work periods", but that they would be able to make choices throughout the experiment that would give them a chance to watch a Youtube video instead. The number of remaining work periods in the experiment was shown at the top of the screen, as was the number of seconds

left in the current work period.

The slider task was based on a task previously used by Gill & Prowse (2012). In each period of the slider task, five horizontal sliders appeared on the screen. Each started at a random setting between 0 and 100, with the slider's value shown to its right, with a random horizontal offset so that the sliders were not aligned. The participant's task was to use the mouse to move each slider to "50" before the work period ended. When a participant released the mouse at the correct setting, the slider turned green to show it had been completed. To ensure that the task took close to the allotted 30 seconds, at the beginning of the task only the top slider was enabled, and the other four were grayed out. Additional sliders were enabled at five-second intervals, such that all five sliders were available after 20 seconds.

Before beginning the experiment, participants chose one of four videos available on Youtube: an episode of "Planet Earth", and episode of "The Great British Bakeoff", and episode of "Mythbusters", or an Ellen Degeneres comedy special. The video was embedded in the experiment with all user controls (such as skipping ahead) disabled. When given access to the video, participants had to keep the browser window open and hold down the space bar for the video to play. This allowed us to ensure that participants maintained engagement with the content.

In the main part of the experiment, participants completed a total of 70 work periods. For the first 10 work periods, participants simply clicked a button to begin the slider task. After these initial periods, participants gained access to six "machines", each of which had a "current spinner" and a "new spinner". Before each work period, participants clicked the circled machine, were shown the machine and made a choice about which spinner to run. Based on their choices, there was a chance the machine would perform the work task for the participant, allowing the participant to watch their chosen video instead.

The machine ran immediately after the choice was made, and affected the next work period, as shown in Figure 1 (left). It then "cooled off" for the following five periods, as choices were made with the other five machines.

At each choice, participants had to select between two options: run the machine's "current spinner" or run a "new spinner". Each machine had two circular spinners with arrows at the top. The current spinner was split into five black and five gold wedges. If a participant chose the current spinner, it spun (immediately or after four periods, depending on the condition) and, if the arrow landed on the gold wedge, the machine worked and the participant could watch the video. Initially, the current spinner's gold wedges were randomly set for each machine to cover between 1/3 and 2/3 of the spinner.

The new spinner initially showed a question mark. If a participant chose the new spinner, then (immediately or after four periods) a new spinner was created and appeared on the machine. In the high-explore condition, the new spinner's gold wedges could cover anywhere from 25% of the spinner

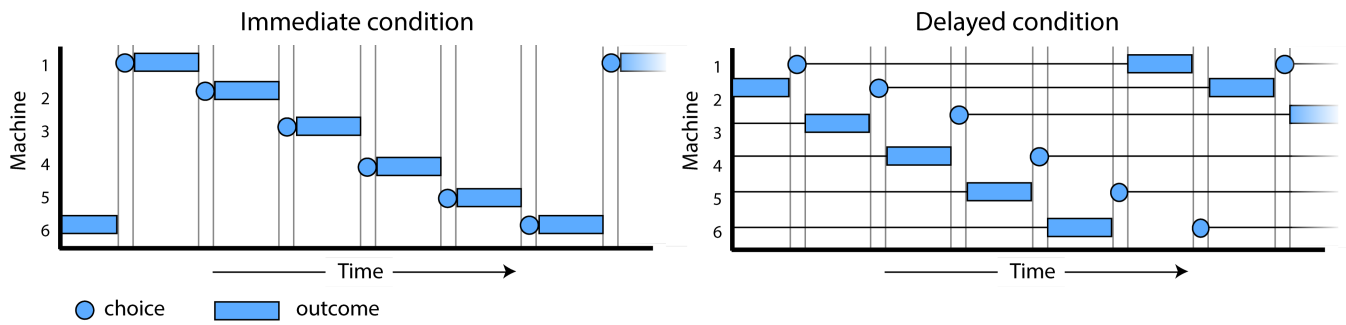


Figure 1: A schematic illustration of the experiment timeline. In all experiments and conditions, six 30-second work periods pass between consecutive choices with a given machine. In Experiment 1 and the immediate condition (left) of Experiment 2, the outcome for a choice is revealed and occurs immediately after the choice is made. In the delayed condition of Experiment 2 (right), the outcome of a choice is revealed and occurs after a delay of four work periods.

to 100% of the spinner. In the low-explore condition, the new spinner's gold wedges could cover anywhere from 0% of the spinner to 75% of the spinner. The new spinner then spun and, if the arrow landed on a gold wedge, the machine worked. Additionally, if the new spinner had a greater gold area than the current spinner, the new spinner was "saved" and the current spinner was updated to the new spinner. This meant that while choosing a new spinner carried risk, it could carry long-term benefits as the current spinner could be improved from its initial value. Importantly, the value of choosing a new spinner was higher in the high-spinner condition, because the range of outcomes was shifted towards higher values.

Finally, in order to induce exploration throughout the entire experiment, the six machines would occasionally "reset" after they ran. When this occurred, the current spinner would be set to a new random value between 1/3 and 2/3 gold. Participants were informed that this would occur randomly on 1/6 of trials. In fact, the procedure was designed so exactly one of the six machines would reset on each cycle through the machines, and no machine would be reset on two consecutive uses.

During the machine choices, participants had access to an "info" button at the bottom of the screen that provided reminders about the dynamics of the task. In addition, before beginning the full experiment, participants completed two practice phase. First, they performed several trials of practice using the machines, with the actual work periods removed. Then, they performed three work periods practicing the slider task.

Participants were given a performance-based bonus of \$3.00 for completing the slider tasks accurately. If they missed fewer than 10% of sliders throughout the experiment and left the video paused less than 20% of the time, they were not penalized. However, if they missed more sliders or left the video paused for longer, they lose 10 cents from their bonus for each additional percentage of sliders missed or time with the video paused. The running percentage of sliders missed and video pause time was displayed at the top of the screen throughout the experiment.

Following the experiment, participants were asked to rate

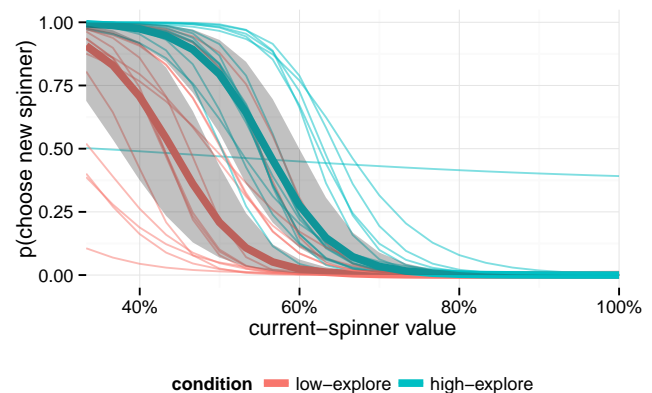


Figure 2: Model-based estimates of participants' probability of choosing a new spinner for different values of the current spinner in Experiment 1. Thick lines and shaded regions indicate the mean and 95% posterior interval for the population-level parameters, while the thin lines indicate the mean posterior parameters for each of the 31 individual participants. Participants in the high-explore condition were more likely to explore at a given current-spinner value than those in the low-explore condition.

their enjoyment of the slider task and of the video-watching task on a 1 to 7 scale.

Results

Participants rated their enjoyment of the video at 5.77 out of 7, on average, and their enjoyment of the slider task at 2.87 out of 7, on average. This difference in enjoyment between the tasks was significant, $t(30) = 10.53$, $p < .001$.

To analyze participants trial-by-trial decision-making behavior, we conducted a hierarchical Bayesian logistic regression using the Stan modeling language (Stan Development Team, 2015). This approach allowed us to estimate population-level effects of the current-spinner value and of condition, while also allowing for individual differences. The regression model included an intercept term as well as terms for the value of the current spinner, the participant's condition, and a condition by value of current spinner interaction.

The results of the regression are shown in Figure 2.

Experiment 2

Method

Participants.

- Abramson, L. Y., Metalsky, G. I., & Alloy, L. B. (1989). Hopelessness Depression: A Theory-Based Subtype of Depression. *Psychological Review*, 96(2), 358–372.
- Abramson, L. Y., Seligman, M. E., & Teasdale, J. D. (1978). Learned helplessness in humans: critique and reformulation. *Journal of abnormal psychology*, 87(1), 49–74. doi: 10.1037/0021-843X.87.1.49
- Bartels, D. M., & Urminsky, O. (2015). To Know and to Care: How Awareness and Valuation of the Future Jointly Shape Consumer Spending. *Journal of Consumer Research*, 41(6), 1469–1485. Retrieved from <http://ezp-prod1.hul.harvard.edu/login?url=http://search.proquest.com/docview/18453927&site=ehost-secure> doi: 10.1086/680670
- Brown, A. L., Chua, Z. E., & Camerer, C. F. (2009). Learning and visceral temptations in dynamic saving experiments. *The Quarterly Journal of Economics*(February), 197–231.
- Bussemeyer, J. R., & Rapoport, A. (1988, jun). Psychological Models of Deferred Decision Making. *Journal of Mathematical Psychology*, 134(2), 91–134. Retrieved from [http://www.sciencedirect.com/science/article/pii/00222496\(88\)90042-9](http://www.sciencedirect.com/science/article/pii/00222496(88)90042-9) doi: 10.1016/0022-2496(88)90042-9
- Denrell, J. (2005). Why most people disapprove of me: experience sampling in impression formation. *Psychological review*, 112(4), 951–978. Retrieved from <http://doi.apa.org/psycinfo/2005-13637-011>
- and choosing in an uncertain world : An investigation of the explore-exploit dilemma in static and dynamic environments. *Cognitive Psychology*, 85, 43–77. Retrieved from <http://dx.doi.org/10.1016/j.cogpsych.2016.01.001> doi: 10.1016/j.cogpsych.2016.01.001
- Rich, A. S., & Gureckis, T. M. (2017). Exploratory choice reflects the future value of information. *Decision*, (in press).
- Sang, K., Todd, P. M., & Goldstone, R. L. (2011). Learning near-optimal search in a minimal explore/exploit task. *Proceedings of the Thirty-third Annual Conference of the Cognitive Science Society*, 2800–2805.
- Shook, N. J., & Fazio, R. H. (2008). Interracial roommate relationships: An experimental field test of the contact hypothesis: Research article. *Psychological Science*, 19(7), 717–723. doi: 10.1111/j.1467-9280.2008.02147.x
- Solnick, J. V., Kannenberg, C. H., Eckerman, D. A., & Waller, M. B. (1980). An experimental analysis of impulsivity and impulse control in humans. *Learning and Motivation*, 11(1), 61–77. doi: 10.1016/0023-9690(80)90021-1
- Stan Development Team. (2015). *Stan: A C++ Library for Probability and Sampling, Version 2.7*. Retrieved from <http://mc-stan.org/>
- Teodorescu, K., & Erev, I. (2014a). Learned Helplessness and Learned Prevalence: Exploring the Causal Relations Among Perceived Controllability, Reward Prevalence, and Exploration.

Psychological Science, 25(10), 1861–1869. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/25193942> <http://pss.sagepub.com/lookup/doi/10.1177/0956797614543022>
doi: 10.1177/0956797614543022

Teodorescu, K., & Erev, I. (2014b). On the Decision to Explore New Alternatives: The Coexistence of Under- and Over-exploration. *Journal of Behavioral Decision Making*, 27, ' . doi: 10.1002/bdm

Tversky, A., & Edwards, W. (1966). Information versus reward in binary choices. *Journal of Experimental Psychology*, 71(5), 680.

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*.

Zwick, R., Rapoport, A., Lo, A. K. C., & Muthukrishnan, a. V. (2003). Consumer Sequential Search: Not Enough or Too Much? *Marketing Science*, 22(4), 503–519. doi: 10.1287/mksc.22.4.503.24909