Drift Rate Reflects Value, Response Bias Reflects Habit: A DDM Analysis of the Reward Pairs Task

Viktor Timokhov* (viktor.timokhov@econ.uzh.ch), Hugo Fluhr (hugo.fluhr@econ.uzh.ch), Philippe Tobler (phil.tobler@econ.uzh.ch), Stephan Nebe (stephan.nebe@econ.uzh.ch)
Zurich Center for Neuroeconomics, Department of Economics, University of Zurich, Switzerland

*: corresponding author

Abstract

Human behavior is guided by both goal-directed and habitual systems. While reinforcement learning (RL) models have captured these influences using choice data, integrating RL with sequential sampling models like the drift-diffusion model (DDM) enables modeling of both choices and response times (RTs). In this study, we tested whether value differences modulate the DDM drift rate and whether prior choice frequency affects response bias simultaneously. Using data of 213 participants in the Reward Pairs Task - an instrumental learning paradigm that independently manipulates stimulus value and choice frequency - we applied hierarchical DDM modeling with collapsing boundaries. Results showed that the best-fitting model captured both value-based and habit-based influences: drift rate scaled with value differences, and response bias reflected differences in choice frequencies. Posterior predictive checks confirmed alignment with observed behavior. These findings support a dual-process view of decision-making, showing that goal-directed and habitual factors influence choice and decision speed via distinct mechanisms.

Keywords: drift-diffusion model; habits; reinforcement learning; cognitive modeling

Introduction

Human behavior is believed to be shaped by two distinct but often interacting systems: a goal-directed system, which selects actions based on the value of expected outcomes, and a habitual system, which elicits responses based purely on stimuli-response associations (Daw, Niv, & Dayan, 2005; Dolan & Dayan, 2013; Miller, Shenhav, & Ludvig, 2019; Huys & Seriès, 2022). Computational models arising from the reinforcement learning (RL) framework have been widely used to explain both habitual and goal-directed behaviors (Daw et al., 2005; Miller, Shenhav, & Ludvig, 2019).

Such models typically focus on predicting choice data, but response times (RTs) offer valuable additional insights (Konovalov & Krajbich, 2019). Integrating RL with sequential sampling models like the drift-diffusion model (DDM) enables joint modeling of choices and RTs, with drift rate reflecting value differences (Pedersen et al., 2017; Fontanesi et al., 2019; Miletić et al., 2020). Fewer studies have explored how habitual learning fits within the sequential sampling framework. Research in perceptual decision-making suggests that choice frequency, whether instructed or learned, can bias the starting point of evidence accumulation toward more frequently chosen options (Leite & Ratcliff, 2011; Mulder et al., 2012; but see Urai et al., 2019). Recent theoretical work (Zhang et al., 2024) proposes that value differences modulate drift rate while habit strength affects response bias - though these assumptions remain largely untested with behavioral data.

The goal of the present study was to validate the difference-based assumptions using behavioral data from the Reward Pairs Task (Nebe, Kretzschmar, Brandt, & Tobler, 2024) - a paradigm that independently manipulates stimulus value for goal-directed action learning and choice frequency for habit formation during training. Previous results suggest that both higher reward value and greater prior choice frequency facilitate faster responding in the Reward Pairs Task (Nebe et al., 2024). We applied hierarchical DDMs (HDDM; Wiecki, Sofer, & Frank., 2013) to examine whether value differences modulate the drift rate, and whether habit strength differences modulate the response bias. Across multiple model specifications, our findings supported this dual influence on decision dynamics.

Methods

Reward Pairs Task. The task (Nebe et al., 2024) is an instrumental learning task designed to independently manipulate reward value and choice frequency. Participants repeatedly choose between pairs of geometric stimuli with fixed reward values (1/3/5/7/9 points), learning these values through feedback across five training sessions. Within each of the intermediate reward levels (3, 5, 7), one stimulus was more often paired with lower-(higher-)value options, increasing (decreasing) its choice frequency during training (Figure 1). Each response was limited to 800ms. A test session followed after five days of training, freely pairing each stimulus with each other and omitting feedback to prevent further learning. Importantly, some test session choices for the first time concerned equal-reward stimuli which differed only in the past choice history, thus controlling for value to assess the impact of frequency on behavior.

Dataset. Data of 213 participants (Nebe et al., 2024) included on average 133 test trials, 35 of which with equal reward. Choice frequency was operationalized as the proportion of stimulus choice over all choices made during training. Only test data were analyzed to isolate the influence of prior learning without ongoing reinforcement.

HDDM. Due to the time pressure in the task, we utilized DDMs with linearly collapsing boundaries. In the models we fitted, drift rate (*v)* and response bias (*z)* were regressed on value difference (indicated as *val*) or choice frequency difference (indicated as *freq*) in various combinations (see Table 1). We indicated standard DDM with linearly collapsing bounds as a baseline model. The upper boundary corresponded to left choices, the lower to right. We used default non-informative priors in the HDDM package. For the group-only model fitting, we sampled 4 chains of 13,500 samples (3,500 burn-in). For hierarchical model fitting (i.e., each parameter is estimated also for individual participants), we sampled 4 chains of 75,000 samples (25,000 burn-in).

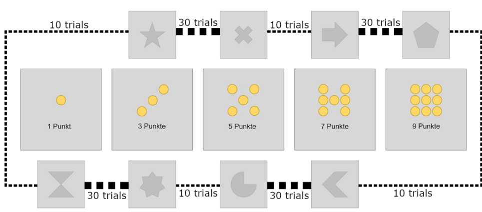

Figure 1: Choice frequency manipulation during training in the Reward Pairs Task

Results

We fitted only group parameters for all combinations of drift rate and response bias with difference-based assumptions. Based on the Deviance Information Criterion (DIC; see Table 1), the second best model (*v_val*) assumes that drift rate is a function of the scaled value difference (as in Miletić et al., 2020). However, the best model (*v_val_z_freq*) additionally assumed that response bias reflects the difference in choice frequencies during training. When fitting the two best models hierarchically, *the v_val_z_freq* model (DIC=-38,059) was still better than the v_val model (DIC=-37,204). All models converged well (R_hat < 1.01) and posterior predictive checks showed closely matched observed choices and RTs in the Reward Pairs Task (Figure 2 for an example). Slope coefficients for both *v_val* and *v_val_z_freq* models were positive, indicating higher drift rate with increasing value difference and shift of the starting point toward the option chosen more often during training.

Table 1. Results of the group-only model fitting.

Model	DIC
v_val_z_freq	-29,081
v_val	-28,562
v_freq_z_val	-23,250
z_val	-21,667
v_freq	-18,520
z_freq	-17,272
Baseline (left VS right)	-14,705

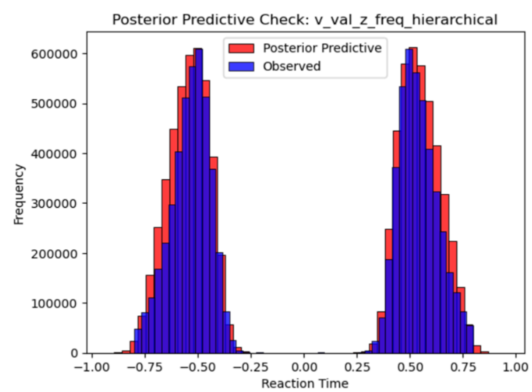

Figure 2: Posterior predictions of the hierarchical *v_val_z_freq* model (positive RT if left stimulus chosen, negative RT if right stimulus chosen,).

Conclusion

We provide empirical support for dissociable mechanisms underlying value-based actions and habits. Specifically, drift rate increased with value differences, while response bias increased with choice frequency differences. These findings guide future modeling of goal-directed and habitual behavior using both choice and response time data. They also suggest that habit and goal-directed systems can interact within a single decision process via distinct parameters (i.e., drift rate and response bias), rather than functioning as separate decision-making modes.

References

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience, 8*(12), 1704-1711. https://doi.org/10.1038/nn1560

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron, 80*(2), 312-325. http://dx.doi.org/10.1016/j.neuron.2013.09.007

Fontanesi, L., Gluth, S., Spektor, M. S., & Rieskamp, J. (2019). A reinforcement learning diffusion decision model for value-based decisions. *Psychonomic Bulletin & Review, 26*(4), 1099–1121. https://doi.org/10.3758/s13423-018-1554-2

Huys, Q.J.M., Seriès, P. (2022). Reward-Based Learning, Model-Based and Model-Free. In: Jaeger, D., Jung, R. (eds) *Encyclopedia of Computational Neuroscience*. Springer, New York, NY. https://doi.org/10.1007/978-1-0716-1006-0_674

Konovalov, A., & Krajbich, I. (2019). Revealed strength of preference: Inference from response times. *Judgment and Decision making, 14*(4), 381-394.

Leite, F. P., & Ratcliff, R. (2011). What cognitive processes drive response biases? A diffusion model analysis. *Judgment and Decision Making, 6*(7), 651–687. https://doi.org/10.1017/S1930297500002680

Miletić, S., Boag, R. J., & Forstmann, B. U. (2020). Mutual benefits: Combining reinforcement learning with sequential sampling models. *Neuropsychologia, 136*, 107261. https://doi.org/10.1016/j.neuropsychologia.2019.107261

Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019). Habits without values. *Psychological Review, 126*(2), 292–311. https://doi.org/10.1037/rev0000120

Mulder, M. J., Wagenmakers, E. J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience, 32*, 2335–2343. https://doi.org/10.1523/JNEUROSCI.4156-11.2012.

Nebe, S., Kretzschmar, A., Brandt, M. C., & Tobler, P. N. (2024). Characterizing human habits in the lab. *Collabra: Psychology, 10*(1). https://doi.org/10.1525/collabra.92949

Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. Psychonomic *Bulletin & Review, 24*(4), 1234–1251. https://doi.org/10.3758/s13423-016-1199-y

Urai, A. E., De Gee, J. W., Tsetsos, K., & Donner, T. H. (2019). Choice history biases subsequent evidence accumulation. *Elife, 8*, e46331. https://doi.org/10.7554/eLife.46331

Wiecki, T., Sofer, I., & Frank, M. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics, 7*. https://www.frontiersin.org/articles/10.3389/fninf.2013.00014

Zhang, C., van Wissen, A., Dotsch, R., Lakens, D., & IJsselsteijn, W. A. (2024). A Sequential Sampling Approach to the Integration of Habits and Goals. *Computational Brain & Behavior.* https://doi.org/10.1007/s42113-024-00199-4