Learning expectations shape initial cognitive control allocation

Javier Masís
Princeton Neuroscience Institute
Princeton University
Princeton, NJ, USA
jmasis@princeton.edu

Sebastian Musslick
Institute for Cognitive Science
Osnabrück University
Osnabrück, Germany
sebastian.musslick@uni-osnabrueck.de

Jonathan D. Cohen
Princeton Neuroscience Institute
Princeton University
Princeton, NJ, USA
jdc@princeton.edu

Abstract

Current models of cognitive control frame its allocation as a process of expected utility maximization. The benefits of a candidate action are weighed against the costs of that control allocation (e.g. opportunity costs). Recent theorizing has found that it is normative to account for the value of learning when determining control allocation. Here, we sought to test whether learning expectations could explain people's initial control allocation in a standard dot-motion perceptual task. We found that participants' initial skill level and learning rate in a first block were able to predict their initial willingness to accumulate evidence in a second block, interpreted as a greater control allocation for the task. Our findings support the hypothesis that agents consider learnability when allocating cognitive control.

Keywords: learning; decision making; cognitive control; drift diffusion model; sequential sampling model

Introduction

Typing technique falls into two categories: the easy way (hunting and pecking), and the hard way (touch typing). Why would anyone ever take the hard way? Because, with enough practice, the hard way will lead to faster typing (Logan, Ulrich, & Lindsey, 2016), a better result in the long term. Several considerations underlie this form of intertemporal choice we face throughout our lives. How long into the future will one be typing, much will one get paid for it, and how quickly can one gain proficiency? Driving these questions are parameters that shape a hidden dynamical dimension of the speed-accuracy tradeoff: more time on task (deliberation time in interrogation paradigms) may be suboptimal in the short term, but optimal in the long term because it allows agents to reach proficiency faster (Masís, Musslick, & Cohen, 2021; Masís, Chapman, Rhee, Cox, & Saxe, 2023; Tsetsos, 2023).

The strategic nature of the choice of how to manage this dynamical speed-accuracy tradeoff suggests there may be control mechanisms that manage such decisions. It has been stipulated that cognitive control allocation adjudicates between motivational factors (e.g. reward) by allocating control according to its expected value (Kool, McGuire, Rosen, & Botvinick, 2010; Kurzban, Duckworth, Kable, & Myers, 2013; Shenhav, Botvinick, & Cohen, 2013). Part of that value is near term rewards that would come from immediate performance, the component of reward that is considered in most models (Musslick, Shenhav, Botvinick, & Cohen, 2015; Musslick et al., 2017; Verguts, Vassena, & Silvetti, 2015; Leng, Yee, Ritz, & Shenhav, 2021). What has been less fully considered is the potential value of increases in future reward that would come from improvements in performance through learning. The most direct test of the allocation of cognitive control in the service of learning comes from a study in which rats were found to strategically manage their learning, trading current rewards for faster learning (Masís et al., 2023). However, to our knowledge, a similar test has yet to be carried out in humans.

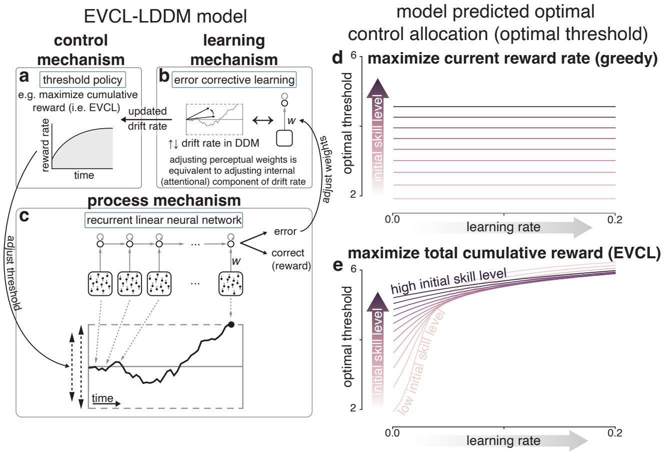

Figure 1: **EVCL-LDDM.** Model is composed of standard decision making and learning components (LDDM, **b** & **c**) with a threshold policy determined by a component that evaluates the expected value of control for learning (EVCL, panel **a**). **(c)** A recurrent linear neural network implements a standard drift diffusion model (DDM). **(b)** The network undergoes error corrective learning during which adjusting the network's perceptual weights is equivalent to adjusting the internal (attentional) component of the drift rate of a DDM. **(a)** A threshold policy controls the evidence accumulation threshold across trials. If that threshold is set to maximize cumulative reward over some time horizon, it is equivalent to an optimal expected value of control for learning (EVCL) model that takes account of the effects of learning, and uses the threshold as the control variable. **(d)** Optimal threshold across learning rate and initial skill level for a greedy policy maximizing current reward rate, and **(e)** a policy maximizing total cumulative reward (EVCL).

Model & Results

Model and Model Predictions

EVCL-LDDM. We combined a recent sequential sampling model, the learning drift-diffusion model (LDDM; Masís et al., 2023), with another recent model that estimaes the expected value of control for learning (EVCL model; Masís et al., 2021) (Fig. 1a-c). LDDM is a process model that imbues the standard drift-diffusion model (DDM) with the ability to learn based on experience. In LDDM, longer deliberation times lead to faster learning because feedback signals are more informative when more stimulus evidence is available to interpret them. As such, slower deliberation can actually be normative. EVCL provides a metacognitive objective to direct LDDM's learning by proposing that agents consider their potential learning trajectory when deciding how much control to allocate to a task.

Predictions. EVCL-LDDM predicts that when maximizing total cumulative reward (EVCL policy), learning expectations (initial skill level and learning rate) determine optimal cognitive control allocation, implemented by adjusting the evidence threshold or, effectively, average deliberation times (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006). Specifically, control should increase with initial skill level and learning rate, and decrease with their interaction (the larger the initial skill level, the smaller the effect of learning rate) (Fig. 1e). In contrast, when maximizing current reward rate (greedy policy), only initial skill level determines optimal control allocation (Fig. 1d).

Study Design and Results

Participants. We collected data using Prolific (`prolific.co`) from 197 participants compensated \$4.80USD (∼\$10.77 an hour) who provided written informed consent in accordance with the relevant Institutional Review Board. After basic engagement exclusions, 159 participants remained.

Study Design. EVCL-LDDM predicts (see Fig. 1) that participants' learning expectations (initial skill level and learning rate) will determine their allocation of control (optimal threshold and therefore also decision times) (Fig. 1e). The model determines optimal thresholds through an unrealistic offline optimization procedure. Instead, people may, after some experience with a task, generate a prior on their learning expectations and use that prior to estimate their optimal control allocation when faced with a sufficiently similar task again.

To test these predictions, we used a classic perceptual decision making task, the random dot kinematogram, with difficult but learnable motion coherence conditions (5%, 10% & 15% based on a pilot study not shown). The study was composed of an initial inducement period (block 1, 200 trials) during which participants generated their learning expectations, followed by a measurement period (first 25 of 200 trials of block 2) where participants' control allocation was measured as a function of those learning expectations.

Analysis. We measured participants' learning expectations via an HDDM regression (Wiecki, Sofer, & Frank, 2013) of drift rate over trials during block 1 (Fig. 2e) where initial skill level and learning rate were operationalized as the intercept (initial drift rate) and regression coefficient of trial (change in drift rate) respectively. We then tested whether learning expectations in block 1 determined initial control allocation in block 2 through an HDDM regression of threshold (Fig. 2f) and a linear mixed effects regression of decision time (Fig. 2g).

Results. First, we qualitatively fit EVCL-LDDM to average performance in block 1 (Fig. 2a, c), and used those parameters to generate optimal decision times based on a greedy policy and one that considers learning (EVCL). The EVCL learning policy qualitatively matched initial decision times in block 2 (Fig. 2b, d), suggesting that on average participants' initial control allocation reflected their learning expectations.

Next, we obtained individual participant estimates on initial drift rate (initial skill level), and change in drift rate (learning rate) with a DDM regression of drift rate over trials fit on performance in block 1. Group-level posteriors showed that participants did learn throughout block 1 (Fig. 2e). We then used individual participant estimates to regress initial threshold in the first 25 trials of block 2 as a function of initial drift rate, change in drift rate and their interaction. We included the inferred final threshold from block 1 to account for the predicted autocorrelation between threshold across blocks. We found that initial threshold depended on initial drift rate, change in drift rate and their interaction, as predicted by our model (Fig. 2f). Because drift rate was 0 during block 2 (coherence of 0%), decision time should be directly proportional to threshold, and thus serve as an additional check of our predictions. A linear mixed effects regression on decision time found significant effects for initial drift rate and change in drift rate, and a trend for their interaction, as predicted by our model (Fig. 2g).

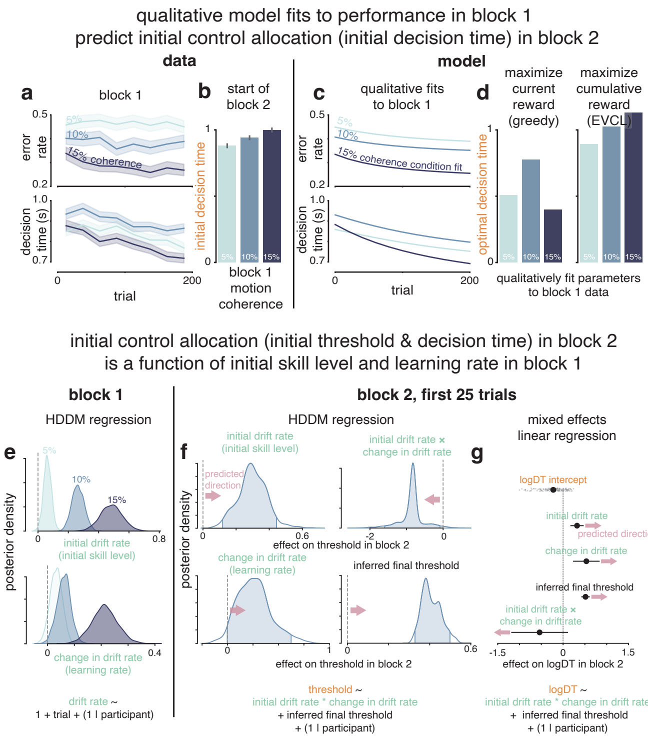

Figure 2: **Learning expectations shape initial control allocation**. **(a)** Mean error rate (ER) and decision time (DT) (25 trial bins, 95% cI) by coherence condition (5%, n = 58; 10%, n = 50; 15%, n = 51). **(b)** mean DT during first 25 trials of block 2, separated by motion coherence condition in block 1. **(c)** ER and DT for qualitative model fits to 5, 10 and 15% coherence conditions. **(d)** Optimal DTs using qualitative parameter fits to block 1 motion coherence data for a greedy policy (left panel) and EVCL policy (right panel). **(e)** Initial drift rate (initial skill level) operationalized as the intercept of drift rate. Change in drift rate (learning rate) operationalized as the coefficient of trial for drift rate (slope of drift rate). **(f, g)** Coefficients for initial drift rate (initial skill level; $p < 0.05$ in **g**), change in drift rate (learning rate; $p < 0.05$ in **g**), initial drift rate×change in drift rate ($p = 0.1$ in **g**), and inferred final threshold ($p < 0.05$ in **g**). 89% high-density intervals shown in **f**. Pink arrows indicate model predictions.

Acknowledgments

J.M. was supported by the Presidential Postdoctoral Research Fellowship at Princeton University, the NIH institutional training grant T32MH065214, and the Swartz Fellowship in Theoretical Neuroscience at Princeton University. S.M. was supported by Schmidt Science Fellows, in partnership with the Rhodes Trust, and the Carney BRAINSTORM program at Brown University. JDC was supported by a Vannevar Bush Faculty Fellowship administered by the Office of Naval Research.

References

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, *113*(4), 700.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665.

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and brain sciences*, *36*(6), 661–679.

Leng, X., Yee, D., Ritz, H., & Shenhav, A. (2021). Dissociable influences of reward and punishment on adaptive cognitive control. *PLoS computational biology*, *17*(12), e1009737.

Logan, G. D., Ulrich, J. E., & Lindsey, D. R. (2016). Different (key) strokes for different folks: How standard and nonstandard typists balance fitts' law and hick's law. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(12), 2084.

Masís, J., Chapman, T., Rhee, J. Y., Cox, D. D., & Saxe, A. M. (2023). Strategically managing learning during perceptual decision making. *Elife*, *12*, e64978.

Masís, J., Musslick, S., & Cohen, J. D. (2021). The value of learning and cognitive control allocation. In *Proceedings of the 43rd Annual Conference of the Cognitive Science Society* (pp. 1837–1843). Vienna, AT.

Musslick, S., Saxe, A., Özcimder, K., Dey, B., Henselman, G., & Cohen, J. D. (2017). Multitasking capability versus learning efficiency in neural network architectures. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 829—834). London, UK.

Musslick, S., Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2015). A computational model of control allocation based on the expected value of control. In *Reinforcement Learning and Decision Making Conference 2015.*

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240.

Tsetsos, K. (2023). Unlocking a new dimension in the speed–accuracy trade-off. *Trends in Cognitive Sciences*.

Verguts, T., Vassena, E., & Silvetti, M. (2015). Adaptive effort investment in cognitive and physical tasks: A neurocomputational model. *Frontiers in Behavioral Neuroscience*, *9*, 57.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). Hddm: Hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in neuroinformatics*, 14.