

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Learning expectations shape initial cognitive control allocation

Permalink

<https://escholarship.org/uc/item/6xd2r257>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Masís, Javier Alejandro

Musslick, Sebastian

Cohen, Jonathan

Publication Date

2024

Peer reviewed

Learning expectations shape initial cognitive control allocation

Javier Masís

Princeton Neuroscience Institute
Princeton University
Washington Rd, Princeton, NJ, USA
jmasis@princeton.edu

Sebastian Musslick

Institute of Cognitive Science
University of Osnabrück
49069 Osnabrück, Germany
sebastian.musslick@uni-osnabrueck.de

Jonathan D. Cohen

Princeton Neuroscience Institute
Princeton University
Washington Rd, 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 subjects' 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 the learnability of a task when deciding how much cognitive control to allocate to that task.

Keywords: learning; decision making; cognitive control; drift diffusion 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, Chapman, et al., 2023). However, to our knowledge, a similar test has yet to be carried out in humans.

Here, we sought to examine whether people allocate cognitive control as a function of their learning expectations. To generate model-based predictions, we combined a recent sequential sampling model, the learning drift-diffusion model (LDDM; Masís, Chapman, et al., 2023), with another model that addresses the expected value of control for learning (EVCL model; Masís et al., 2021). The 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 there is more stimulus evidence available to interpret them. As such, slower reaction times can actually be normative. The EVCL provides a metacognitive objective to direct LDDM's learning by proposing that agents consider their own potential learning trajectories when deciding how much control to allocate to a particular task. The combined EVCL-LDDM model predicted that learning expectations (composed of initial skill level, learning rate and their interaction) determined optimal cognitive control allocation (implemented by adjusting the evidence accumulation threshold or, effectively, average deliberation times, to optimize reward rate; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006). To test these predictions, we used a classic perceptual decision making task, the random dot kinematogram, the difficulty of which was manipulated by the motion coherence of a moving dot stimulus. In a first block, participants completed a difficult, but learnable set of trials with the aim of inducing a set of learning expectations. In a second block, we measured participants' early evidence thresholds and deliberation times. We found that participants' performance in the first block (i.e., their initial skill levels and learning rates) predicted their early evidence thresholds and deliberation times in the second block. These results suggest that people allocate cognitive control as a function of their learning expectations.

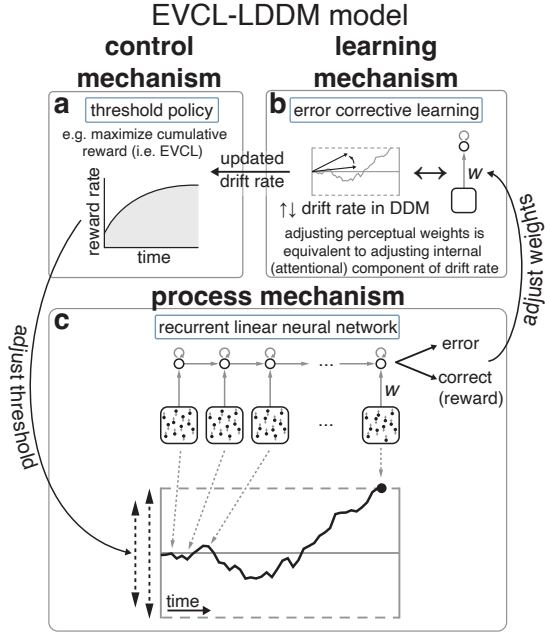


Figure 1: **Graphical description of EVCL-LDDM model.** Model contains a standard decision making and learning component (LDDM, **b** & **c**) and a component that sets a threshold policy by evaluating the expected value of control for learning (EVCL, **a**). (**c**) A recurrent linear neural network implements a standard DDM. (**b**) Network undergoes error corrective learning. Adjusting network’s weights is equivalent to adjusting attentional component of the drift rate of a DDM. (**a**) A threshold policy controls the evidence accumulation threshold across trials. If the threshold is set to maximize cumulative reward over some time horizon, it is equivalent to an optimal EVCL model that accounts for the effects of learning, and uses threshold as the control variable.

EVCL-LDDM Model

We formulated a model with three required elements: 1) a process mechanism to make choices; 2) a learning mechanism to improve choices with experience; and 3) a control mechanism that specifies some objective, such as maximizing total cumulative reward, that it will seek to optimize while taking into account its learning expectations. The EVCL-LDDM model meets these requirements by nesting two previous models working at different levels of abstraction. First, the learning drift-diffusion model (LDDM; Masís, Chapman, et al., 2023) expands the DDM, a process model for making choices (Ratcliff & Rouder, 1998), with a mechanism that allows learning through experience. Second, the expected value of control for learning model (EVCL; Masís et al., 2021) provides a control mechanism that can be added to LDDM in the form of an objective, such as the maximization of total cumulative reward, in order to shape its learning prospects and performance.

Learning drift-diffusion model (LDDM)

The DDM is often implemented as a static model without time-varying parameters. However, learning is an inherently dynamic process, and a model of learning should have a mechanism for modifying its parameters over time. For example, for a stimulus of constant difficulty, learning can be represented as an increase in drift rate. The LDDM (Masís, Chapman, et al., 2023) takes this approach, implementing a DDM in the form of a simple linear recurrent neural network model, in which changes in drift occur through changes in connection strengths as a function of a standard error-driven (backpropagation) learning algorithm. Furthermore, the model has an analytical solution for the average learning dynamics that can be used to determine the parameters that optimize a specified objective or control policy, without requiring computationally-intensive simulations.

Expected Value of Control for Learning (EVCL)

Threshold is a critical factor that determines how much integration takes place. In previous work, it has been shown that, for a given set of parameters, and assuming a fixed drift rate, there is a single optimal threshold that maximizes reward rate (Bogacz et al., 2006). However, to the extent that drift rate changes with learning, it merits considering how people adapt their threshold to take account of this. Furthermore, to the extent that threshold controls the amount of integration (and thereby accuracy of the response) on each trial, a higher threshold may lead to more effective learning. This observation suggests that in settings where the agent anticipates it may be able to learn, it may be optimal to set a higher than optimal threshold initially in order to promote integration and learning that may compensate by yielding better performance and therefore higher rewards in the future.

If we consider threshold setting to be a cognitive control process, then threshold choice can be modeled with a normative theoretical model for the allocation of cognitive control, the EVCL model (Masís et al., 2021), an extension of the well-supported opportunity-cost based model of control, the Expected Value of Control theory (EVC; Shenhav et al., 2013), in which the agent considers the impact of its control choices on its future self in the form of its skill level or automaticity and the corresponding impact on the agent’s reward. Following Masís et al. (2021), the expected value of control for a particular control signal in a particular state is given by the difference between the expected payoff of control in that state and the cost of that control signal

$$\text{EVC}(\text{signal}, \text{state}) = \mathbf{E}[\text{Payoff}(\text{signal}, \text{state})] - \text{Cost}(\text{signal}) \quad (1)$$

Here we use threshold in the LDDM as the control signal. The agent’s current skill level, error rate and decision time, and completed trials thus far comprise its state. The expected payoff is computed as the expected value of possible outcomes (correct and incorrect, indexed by i) given the control signal and state, weighted by their respective probabilities

$$E[\text{Payoff}(\text{signal}, \text{state})] = \sum_i P(\text{outcome}_i | \text{signal}, \text{state}) \cdot \text{Value}(\text{outcome}_i) \quad (2)$$

The value function is, in turn, comprised of two elements

$$\text{Value}(\text{outcome}) = R_0(\text{outcome}) + \gamma \cdot \max_j [\text{EVC}(\text{signal}_j, \text{outcome})] \quad (3)$$

The first (R_0) is the immediate reward for the current outcome, and the second ($\gamma \cdot \max_j [\text{EVC}]$, where j indexes the possible control signals, i.e., possible evidence thresholds) is the future discounted reward for the control signal that yields the greatest expected reward when the outcome of the current state is used as the next state. The discount factor γ controls whether the model is fully myopic ($\gamma = 0$) or forward-looking with no discounting ($\gamma = 1$).

We define the immediate reward R_0 in our model (eq. 3) as the instantaneous or current reward rate iRR

$$R_0 \equiv \text{iRR} = \frac{1 - \text{ER} - q\text{ER}}{\text{DT} + D_{\text{tot}}} \quad (4)$$

where ER is error rate, DT is decision time, and D_{tot} captures the response-to-stimulus intervals and the non-decision component of reaction time, and we allow for a reward penalty q for errors, referred to as an accuracy bias (Zacksenhouse, Bogacz, & Holmes, 2010; Bogacz, Hu, Holmes, & Cohen, 2010; Balci et al., 2011). Further, because the cost of a high threshold is implicitly included in the agent’s reward rate payoff (a higher threshold would lead to a higher average decision time, and thus a lower reward rate), we do not include an explicit cost to the agent’s threshold choice (eq. 1).

Because the value function uses outcome of the current state (eq. 4) to compute the EVC for the next state (see eq. 3), EVCL can be used to recursively simulate the consequences of its control choices over a reasonable range of control signals (threshold choices) and up to some tractable future time horizon. The optimal control signal signal^* (threshold choice) for the current state is then chosen by selecting the control signal with the maximum EVC.

$$\text{signal}^* \leftarrow \operatorname{argmax}_i [\text{EVC}(\text{signal}_i, \text{state})] \quad (5)$$

Threshold Control Policies

As a benchmark, we compare the EVCL policy (that maximizes the integral of reward over some predetermined time window) with a simple “greedy” policy (the maximization of instantaneous or current reward rate, Gold & Shadlen, 2002). For the greedy policy, we set the discount factor $\gamma = 0$ in eq. 3 to make the model fully myopic, and thus choose the threshold that maximizes current reward rate. For the EVCL policy, we set the discount factor $\gamma = 1$ to make the model value

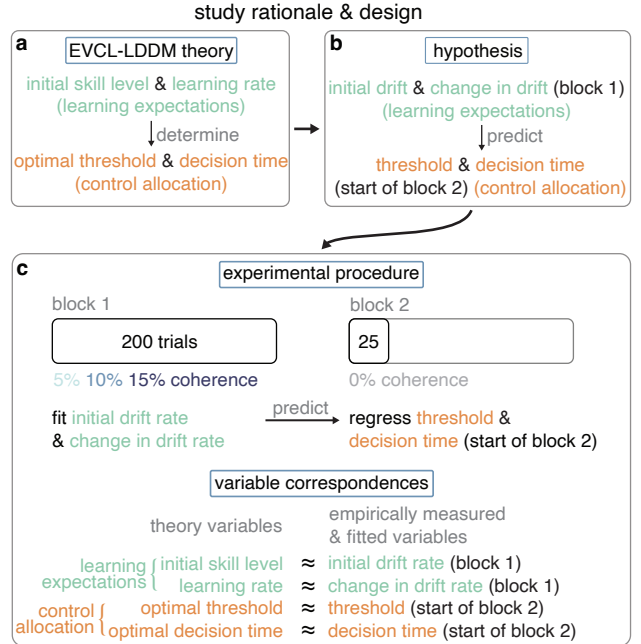


Figure 2: Study rationale & design. (a) The EVCL-LDDM model predicts that initial skill level & estimated learning rate (learning expectations) are used to determine the optimal threshold (control allocation) and thereby decision time (see Fig. 3e). (b) Thus, we hypothesize that initial drift rate & change in drift rate (used to estimate learning rate) during block 1 should predict initial threshold (control allocation) & decision time at the start of block 2. (c) To test our hypothesis, we will fit the initial drift rate & change in drift rate over block 1 and test the extent to which these predict initial threshold and decision time during the start (first 25 trials) of block 2.

rewards at the end of the horizon as much as immediate rewards, and thus choose the threshold that maximizes total cumulative reward.

For tractability in threshold optimization, while forward looking simulations sampled different possible thresholds, threshold was kept fixed over the duration of each simulation (i.e., over its temporal horizon). Finally, because we are only interested in evaluating the impact of learning expectations on early threshold setting, we report the optimal threshold at the beginning of the horizon.

Study Description

Rationale

The EVCL-LDDM model predicts (see Fig. 3e) that participants’ learning expectations (learning rate and initial skill level) will determine their allocation of control (optimal threshold and therefore also decision times) (Fig. 2a). The model determines optimal thresholds through an offline optimization procedure, which is likely not what people do. Instead, people may, after some experience with a task, generate

a prior on their learning expectations for that task and use that prior to estimate their optimal control allocation when faced with a sufficiently similar task again.

To test this prediction, we sought to design a study with two components: an inducement period, during which participants could generate their learning expectations, followed by a measurement period, during which participants' control allocation could be measured as a function of those learning expectations. Making use of a task that could be modeled using the DDM, we hypothesized that subjects' learning expectations (learning rate and initial skill level), operationalized as initial drift rate and change in drift rate shaped during a first inducement block, should predict their cognitive control allocation, reflected in their choice of early threshold and decision time during a second measurement block (Fig. 2b).

Design

To test this hypothesis, we designed a study composed of two blocks of equal length (200 trials) of random dot motion (Fig. 2c). During block 1, the inducement period, participants were presented with motion coherences of 5%, 10% or 15%. Pilot studies that included confidence judgments and surveys (not shown here) indicated that in this coherence range participants reliably learned but were not reliably aware of their learning. Operating just beyond participants' awareness of learning would allow us, we reasoned, to measure control allocation while reducing the interference of overt strategies. During block 2, the measurement period, participants were presented—unawares—with a motion coherence of 0%, i.e., random noise, in a direction orthogonal to what they saw in block 1. The change in motion direction and the equal block length (despite the fact that we would only consider early trials in our analysis) were chosen to communicate that block 2 contained a very similar but distinct task to block 1. Presenting participants with random noise in block 2 served a twofold purpose. First, it would allow us to measure participants' choice of early thresholds (e.g., first 25 trials) based on—we hypothesized—learning expectations (or priors) formed during block 1 before they were updated with new evidence from block 2. Second, it would allow us to use decision time as a secondary measure of control allocation: a motion coherence of 0% all but guarantees a drift rate of 0, which means that decision time depends entirely on threshold choice. We chose 25 trials in an attempt to balance our experimental desire for few early trials with the reliable and accurate recovery of latent participant parameters. Figure 2c includes a summary of our theoretical and empirically measured and fitted variable correspondences.

Participants We collected data online from 197 participants, each receiving \$4.80USD (~\$10.77 per hour), using Prolific (prolific.co). Participants provided written informed consent in accordance with the relevant Institutional Review Board. After basic engagement exclusions, 159 participants remained. Of these, 58, 50, and 51 performed the 5, 10 and 15% coherence conditions during block 1.

Analysis During block 1, we regressed drift rate and threshold on trial to estimate their evolution over the block.

$$\text{drift rate / threshold} \sim 1 + \text{trial} + (1 + \text{trial}|\text{participant}) \quad (6)$$

Participant drift rate intercepts (initial drift rate) and slopes (change in drift rate) were used as measures of initial skill level and learning rate. To account for and measure effects above and beyond the potential autocorrelation of threshold in block 1 and 2, we computed an inferred final threshold for each participant with their threshold intercept and slope estimates to be included in the regressions for block 2.

During the first 25 trials of block 2, we regressed threshold on initial drift rate, change in drift rate, their interaction, and inferred final threshold from block 1.

$$\text{threshold} \sim 1 + \text{initial drift rate} * \text{change in drift rate} + \text{inferred final threshold} + (1|\text{participant}) \quad (7)$$

Additionally, we regressed the log of mean decision time during the first 25 trials of block 2 with the same predictors using mixed effects linear regression.

$$\log \text{DT} \sim 1 + \text{initial drift rate} * \text{change in drift rate} + \text{inferred final threshold} + (1|\text{participant}) \quad (8)$$

We fit DDM regressions to the data using the HDDM-LAN(nn) extension (Fengler, Govindarajan, Chen, & Frank, 2021) of the Hierarchical Drift Diffusion Model (HDDM) package in Python (Wiecki, Sofer, & Frank, 2013). Mixed effects linear regressions were carried out with the pymer4 package for Python (Jolly, 2018).

Results

Model Predictions

We computed optimal thresholds under two contrasting objectives: a greedy policy that maximized current reward rate, and an EVCL policy that maximized total cumulative reward over a given time horizon. We first qualitatively fit the base model to the behavioral data (Fig. 4), and then used those parameters to simulate the model under each objective. Figure 3a-c shows a simple example of the optimization procedure.

A greedy policy (myopic) will not consider the learning trajectory. As such, the agent's learning rate is irrelevant (Fig. 3d). In contrast, for the range of parameters explored (see **Discussion**), the EVCL policy makes three distinct predictions (Fig. 3e). First, optimal threshold increases as a function of initial skill level to perform the task. Second, optimal threshold increases as a function of learning expectation (predicted learning rate). Third, the effect of learning expectation on optimal threshold is mediated by initial skill level: the greater the initial skill level, the smaller the effect of learning expectation.

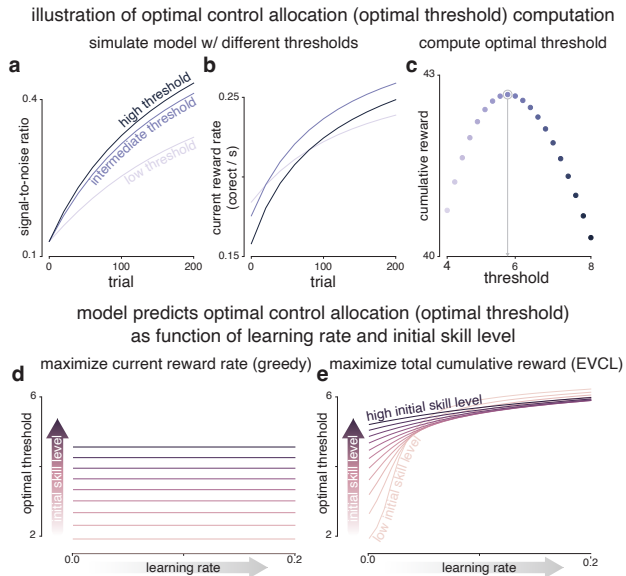


Figure 3: EVCL-LDDM model predictions for optimal control allocation (optimal threshold) as a function of initial skill level and learning rate. In order to compute the optimal threshold for some policy (i.e., maximize total cumulative reward) for a given set of parameters (initial skill level and fixed learning rate), we simulate the model with different thresholds, and visualize the (a) signal-to-noise ratio (SNR, drift/noise), (b) current reward rate (correct / second) and (c) cumulative reward and select the threshold that maximizes cumulative reward. (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).

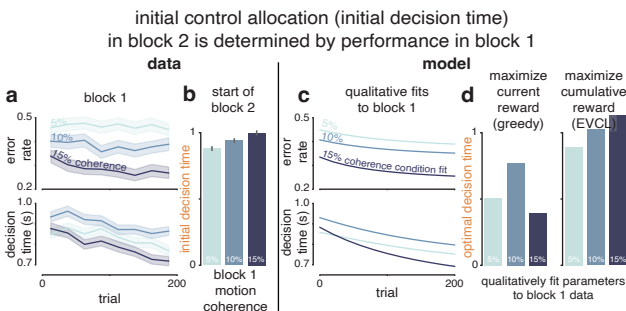


Figure 4: Initial decision time in block 2 determined by performance in block 1. (a) Mean error rate and decision time (25 trial bins, 95% confidence interval) by coherence condition (5%, $n = 58$; 10%, $n = 50$; 15%, $n = 51$). (b) mean decision time during first 25 trials of block 2, separated by motion coherence condition in block 1. (c) Error rate and decision time for qualitative model fits to 5, 10 and 15% coherence conditions. (d) Optimal decision times using qualitative parameter fits to block 1 motion coherence data for a greedy policy (left panel) and EVCL policy (right panel).

initial skill level & learning rate scale with difficulty
block 1 (group fits)

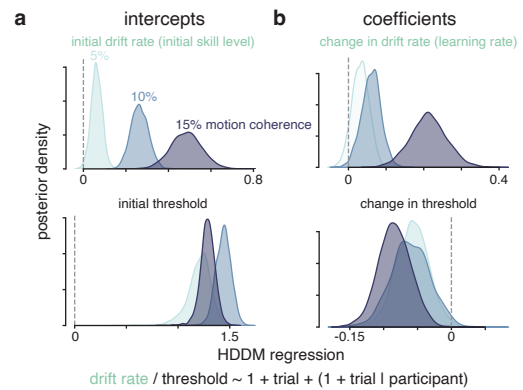


Figure 5: Initial skill level & learning rate scale with difficulty. (a) Initial drift rate (initial skill level) is operationalized as the intercept of drift rate. (b) Change in drift rate (learning rate) is operationalized as the coefficient of trial for drift rate (slope of drift rate). Threshold was allowed to vary (bottom panels, a & b) and generally decreased over trials.

Initial skill level & learning rate scale with difficulty

Group-level DDM regression posterior estimates revealed that, as expected, both initial drift rate (initial skill level) and change in drift rate (learning rate) increased as functions of stimulus difficulty (Fig. 5a & b, top panels), indicating that our aim of inducing a learning prior based on learning outcomes was effective. Group-level posterior estimates for threshold indicated that initial threshold values largely overlapped across coherence, with a trend for a larger threshold for the 10% coherence condition (5a, bottom panel.) Thresholds decreased over the experiment, with a trend towards a larger decrease for the easiest 15% coherence condition (5b, bottom panel). It is not uncommon for thresholds to decrease over the course of a simple perceptual task, which may reflect factors such as boredom and fatigue not considered here.

Learning expectations determine initial control allocation

Qualitative model fits to block 1 used to generate optimal decision times for block 2 suggested participants' improvement (learning) in block 1 determined their initial decision time (control) in block 2 (an EVCL policy; Fig. 4). We next sought to quantitatively test the prediction that initial drift rate (initial skill level), change in drift rate (learning rate) and their interaction during block 1 would all determine initial threshold setting (control allocation) at the start of block 2, beyond the expected autocorrelation with threshold at the end of block 1 (see **Analysis**; Fig. 6b, bottom right panel).

We found that initial drift rate and change in drift rate both had positive effects on threshold (Fig. 6b, left middle and left bottom panels), and the interaction of initial drift rate and change in drift rate had a negative effect on threshold (Fig.

initial control allocation (threshold / decision time) in block 2 as a function of initial skill level and learning rate in block 1
block 2, first 25 trials (group fits)

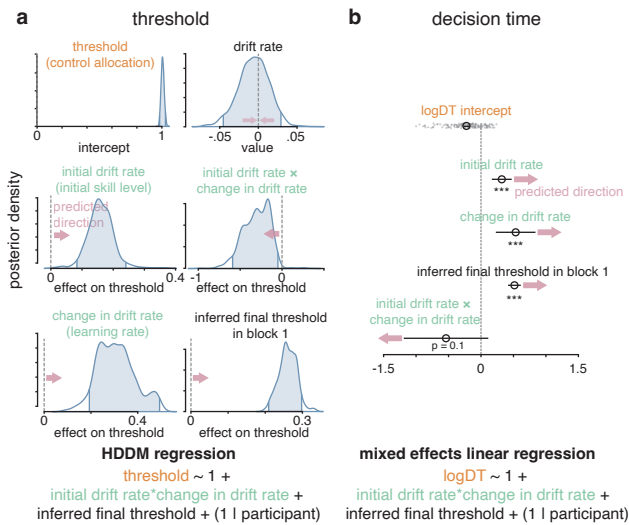


Figure 6: **Initial control allocation (initial threshold / decision time) as a function of initial skill level and learning rate.** (a) HDDM regression and (b) mixed effects linear regression to predict the threshold and decision time in block 2 (first 25 trials) using results from block 1 DDM regression (Fig. 5). Drift rate expected to be 0 because of 0% coherence. Coefficients for initial drift rate (initial skill level), change in drift rate (learning rate), initial drift rate \times change in drift rate, and inferred final threshold all have expected signs based on theory in both regressions. Pink arrows: model predictions. Shaded regions: 89% high-density intervals. ***: $p \leq 0.001$.

6b, middle right panel). These results are consistent with the predictions made by the EVCL threshold policy that maximizes cumulative reward, and not with the greedy threshold policy that maximizes current reward (Fig. 3d & e).

As part of our experimental design, subjects saw 0% motion coherence in block 2 in order to measure threshold as a function of learning expectations set in block 1. We verified this experimental manipulation by finding that the posterior density of drift rate was centered around 0 (Fig. 6a, top right panel). Because drift rate was 0 during block 2, decision time depended solely on threshold choice, and as such served as an additional proxy measure for control allocation. A mixed effects linear regression to test the effects of initial drift rate (initial skill level; $p \leq 0.001$), change in drift rate (learning rate; $p = 0.001$) and their interaction ($p = 0.1$) during block 1 on decision time at the start of block 2 yielded results consistent with model predictions (Fig. 6b).

Discussion

The work here suggests that people selectively allocate cognitive control in response to their learning expectations. We implemented a computational model (EVCL-LDDM) that combines a robust sequential sampling model that can learn

(LDDM) with the normative objective of cumulative reward maximization from a cognitive control allocation model (EVCL) and found that experimental results adhered to the model's predictions, and not to those based on a simpler normative objective (instantaneous reward rate maximization). We reported evidence for learning expectations shaping mental effort allocation in a relatively short time span, indicating that people's biases for learning may form at multiple timescales. Most importantly, our results suggest that people allocate their mental resources in order to modify their own cognitive bounds, as expressed in a recent account of bounded optimality and rationality (Musslick & Masís, 2023).

The notion of adaptive control in humans has been explored in other settings, particularly reinforcement learning, sharing principles with our study. One prominent research effort has investigated whether people explore randomly (e.g., with an information bonus) or strategically (e.g., prioritizing exploration when it has greater potential to improve future rewards, such as when time horizons are longer) (Wilson, Geana, White, Ludvig, & Cohen, 2014). From this effort, work on information seeking (a type of learning) found that people traded immediate reward (exploitation) for information (exploration) when information had more potential to help (Geana, Wilson, Daw, & Cohen, 2016), while work grounded in control theory found that people strategically weighed the costs and benefits when deciding whether to explore unknown actions (e.g., a jam session with friends is better suited for trying silly new sounds on one's musical instrument than a recital) (Schulz, Klenke, Bramley, & Spekenbrink, 2017). Overall, people do seem to explore strategically, that is, explore when they estimate greater benefits and fewer costs to that exploration. This general finding is in line with the theory we have tested here: the decision to allocate cognitive control in the service of learning should consider the extent to which learning is possible and useful.

Future work may wish to explore other tasks (perceptual and value-based) and broader ranges of difficulties than those explored here. Such work should consider that the range of parameters can affect the direction of the relationships between learning expectations and control allocation. For example, with a sufficiently high skill level, a large threshold is not optimal, and thus this relationship is not strictly monotonically increasing. In this study, we utilized difficult stimuli for which these relationships were relatively simple and testable with linear regressions. Future work wishing to explore these relationships across a broader range of difficulties should consider potential non-monotonicities and use appropriate methods to account for them (e.g., quadratic regressions).

Future work may also wish to explore whether the lower-level decision-making process (assumed to be the DDM in this study) may itself have mechanisms that yield effective learning without the explicit need for a metacognitive controller, as suggested by recent work proposing that choices are made when a mix of information and reward per unit time is maximized (Masís, Melnikoff, Barrett, & Cohen, 2023).

Acknowledgments

We would like to thank Harrison Ritz for useful discussions on DDM fitting intricacies. We are grateful to our reviewers for providing thoughtful, encouraging and constructive feedback. J.M. was supported by the Presidential Postdoctoral Research Fellowship at Princeton University, by the NIH institutional training grant T32MH065214, and the Swartz Foundation Fellowship for Theory in 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. J.D.C. was supported by a Vannevar Bush Faculty Fellowship administered by the Office of Naval Research.

References

- Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., & Cohen, J. D. (2011). Acquisition of decision making criteria: reward rate ultimately beats accuracy. *Attention, Perception, & Psychophysics*, *73*, 640–657.
- 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.
- Bogacz, R., Hu, P. T., Holmes, P. J., & Cohen, J. D. (2010). Do humans produce the speed–accuracy trade-off that maximizes reward rate? *Quarterly journal of experimental psychology*, *63*(5), 863–891.
- Fengler, A., Govindarajan, L. N., Chen, T., & Frank, M. J. (2021). Likelihood approximation networks (lans) for fast inference of simulation models in cognitive neuroscience. *Elife*, *10*, e65074.
- Geana, A., Wilson, R. C., Daw, N., & Cohen, J. D. (2016). Information-seeking, learning and the marginal value theorem: a normative approach to adaptive exploration. In *Proceedings of the 38th annual meeting of the cognitive science society* (pp. 1793–1798).
- Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, *36*(2), 299–308.
- Jolly, E. (2018). Pym4: Connecting r and python for linear mixed modeling. *Journal of Open Source Software*, *3*(31), 862. doi: 10.21105/joss.00862
- 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., Melnikoff, D., Barrett, L. F., & Cohen, J. (2023). When to choose: Information seeking in the speed–accuracy tradeoff. In *Proceedings of the 45th annual meeting of the cognitive science society*. Sydney, Australia.
- Masís, J., Musslick, S., & Cohen, J. D. (2021). The value of learning and cognitive control allocation. In *Proceedings of the 43rd annual meeting of the cognitive science society* (pp. 1837–1843). Vienna, Austria.
- Musslick, S., & Masís, J. (2023). Pushing the bounds of bounded optimality and rationality. *Cognitive Science*, *47*(4), e13259.
- 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*.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological science*, *9*(5), 347–356.
- Schulz, E., Klenske, E. D., Bramley, N. R., & Speekenbrink, M. (2017). Strategic exploration in human adaptive control. In *Proceedings of the 39th annual meeting of the cognitive science society*.
- 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*.
- 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.
- Zackhouse, M., Bogacz, R., & Holmes, P. (2010). Robust versus optimal strategies for two-alternative forced choice tasks. *Journal of mathematical psychology*, *54*(2), 230–246.