Title

The Value of Learning and Cognitive Control Allocation

Authors

Masís, Javier Alejandro
Musslick, Sebastian
Cohen, Jonathan

Publication Date

2021

Peer reviewed

The Value of Learning and Cognitive Control Allocation

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

Sebastian Musslick
Princeton Neuroscience Institute
Princeton University
Washington Rd, Princeton, NJ, USA
musslick@princeton.edu

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

Abstract

Current models of cognitive control address selection among tasks in terms of a cost-benefit tradeoff. Importantly, they usually assume a fixed level of competence for each candidate task when estimating its value. However, performing a task can improve competence through learning, which should be factored into estimates of future value. Here, we consider an extension of the Expected Value of Control (EVC) theory that includes such estimates. We demonstrate that control allocation is a function of task learnability, and show the use of this model by generating novel predictions in cognitive effort discounting tasks. We argue that the value of learning in control allocation may account for the seemingly paradoxical finding that sometimes participants prefer more difficult (*i.e.* costly) tasks, and discuss how the model can be leveraged to further our understanding of human decision making and cognitive impairments.

Keywords: learning; decision making; cognitive control; expected value of control theory

Introduction

To achieve long-term goals, humans must constantly adapt their information processing toward relevant tasks. Cognitive control specifies the collection of mechanisms enabling such flexible reconfiguration. A growing number of theories suggest that exerting cognitive control is associated with a cost, and that participants consider this cost when deciding how much control to allocate among tasks (Kool & Botvinick, 2018; Kurzban, Duckworth, Kable, & Myers, 2013; Shenhav, Botvinick, & Cohen, 2013). Such theories successfully explain human behavior in cognitive studies in which participants are asked to choose between tasks of different cognitive demand (Kool, McGuire, Rosen, & Botvinick, 2010). For instance, in the Cognitive Effort Discounting (COGED) Task, participants deciding between an easy task for a low reward and a difficult task for a higher reward often select the easier task, even if it means forgoing a higher reward (Westbrook & Braver, 2015).

In contrast to this proposition, there is mounting evidence that participants sometimes prefer more over less control-demanding tasks despite equal rewards (Cacioppo & Petty, 1982). This “paradox of effort” has led researchers to suggest that exerting cognitive control is intrinsically valuable (Inzlicht, Shenhav, & Olivola, 2018). Yet theories of control allocation lack a normative rationale for an intrinsic value of control, other than the prospect of immediate rewards.

In some situations, this paradox may be a result of the value of learning. Consider the following dilemma. A student must decide whether to continue typing with their index fingers, or to learn to type properly. The meta-decision to learn to type must take into account the predicted future benefits of typing properly because the student must incur the cost of a loss of productivity while learning. In a situation such as this, not only would the student have to choose the more effortful task, but doing so would lead to fewer present rewards than applying the far easier policy of “hunting and pecking.” This dilemma was recently explored in rats and simulated agents, and both rats and agents choose the more effortful and less presently rewarding task of learning in order to improve future rewards (Masís, Chapman, Rhee, Cox, & Saxe, 2020). These results suggest that in some situations, the application of cognitive control in the absence of obvious immediate rewards might be explained by the future discounted value of learning. Nonetheless, little attention has been given to the link between control allocation and the value of learning.[1]

Here, we extend a rational model of control allocation, the Expected Value of Control (EVC) theory (Shenhav et al., 2013), to account for the future value of learning. In former computational implementations, EVC theory has only taken into account instantaneous expected reward from control allocation (Musslick, Shenhav, Botvinick, & Cohen, 2015; Musslick, Cohen, & Shenhav, 2019). However, the theory can be extended to account for the future value of learning, that may be an important but as yet under-addressed component that contributes to the intrinsic value of effort. We demonstrate that taking into account learning during control allocation can lead agents to accrue higher amounts of rewards with less control over the longer term. Further, we derive predictions from this model that can be tested in an extended version of the COGED task. Finally, we discuss how insights from this study help to close current gaps between empirical phenomena and existing models of control allocation, and discuss their role in furthering a comprehensive understanding of cognition.

Learning Expected Value of Control Theory

The Learning Expected Value of Control (LEVC) theory is an extension of the Expected Value of Control (EVC) theory that accounts for the value of learning when allocating cognitive control (Shenhav et al., 2013). We will first describe EVC

[1] *See* Sagiv, Musslick, Niv, and Cohen (2020) & Ravi, Musslick, Hamin, Willke, and Cohen (2020) for related work on the tradeoff between learning efficiency and multitasking ability.

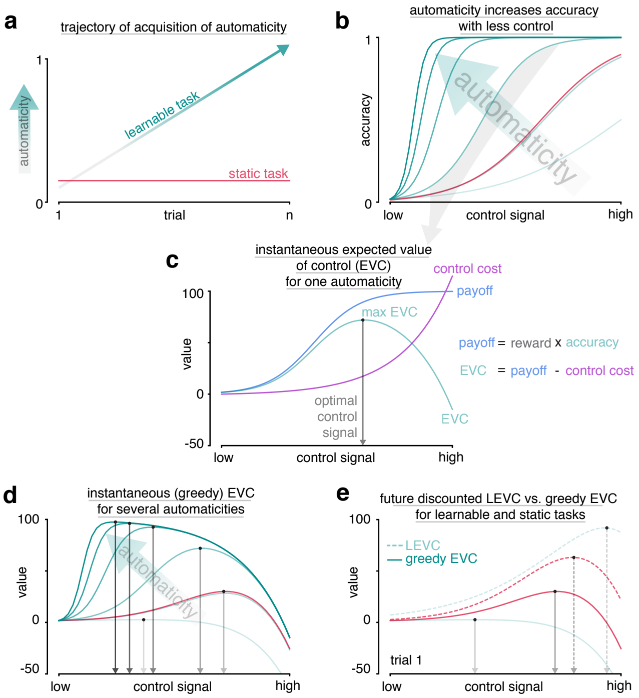

Figure 1: **The value of learning as a determinant of control allocation.** **(a)** Two theoretical tasks, one linearly learnable, and the other static. **(b)** Accuracy is a function of both automaticity and control signal. Higher automaticities result in equal accuracies for less control. **(c)** The instantaneous EVC is the difference between the payoff and cost across feasible control signals. The optimal control signal corresponds to the maximum of the EVC curve. **(d)** Increasing automaticity leads to higher instantaneous EVCs. **(e)** Computing an instantaneous 'greedy' EVC and a future discounted 'learning' EVC leads to different EVCs and optimal control signals, in this example flipping which task has a higher EVC.

theory followed by the LEVC extension.

EVC Theory

In EVC theory, an agent chooses how much cognitive control to exert and where to allocate it based on the expected value of that control (Shenhav et al., 2013). The EVC, calculated for every feasible control signal and agent state, is the difference between the expected payoff and cost of control for that particular combination of state and control signal:

$$\begin{aligned}&\mathrm{EVC}(\mathrm{signal},\mathrm{state}) = \\ &\qquad \mathbf{E}[\mathrm{Payoff}(\mathrm{signal},\mathrm{state})] - \mathrm{Cost}(\mathrm{signal}) \quad (1)\end{aligned}$$

An agent's state can encompass many aspects, such as the particular task being considered, its difficulty, and the agent's (evolving) ability to perform that task (the focus of LEVC theory). Control signal refers to the amount of control allocated, ranging from nearly none (watching television) to substantial (solving a complex equation). The cost of control, frequently modeled as a monotonic function of the control signal, is a reflection of the inherent cost of control (Kool et al., 2010). Here, we model it as an exponential function of the control signal and a cost parameter c:

$$\mathrm{Cost}(\mathrm{signal}) = e^{c\cdot\mathrm{signal}} - 1 \quad (2)$$

The expected payoff is a function of the probability of certain outcomes given a particular control signal and agent state multiplied by the value associated with those outcomes:

$$\begin{aligned}&\mathbf{E}[\mathrm{Payoff}(\mathrm{signal},\mathrm{state})] = \\ &\qquad \sum_i P(\mathrm{outcome}_i|\mathrm{signal},\mathrm{state}) \cdot \mathrm{Value}(\mathrm{outcome}_i) \quad (3)\end{aligned}$$

For example, when an agent is performing a task that it can get correct or incorrect and only correct is rewarded, the expected payoff would be written as follows:

$$\begin{aligned}&\mathbf{E}[\mathrm{Payoff}(\mathrm{signal},\mathrm{state})] = \\ &\qquad P(\mathrm{correct}|\mathrm{signal},\mathrm{state}) \cdot \mathrm{Value}(\mathrm{correct}) + 0\end{aligned}$$

The Value function for a particular outcome has two elements, an immediate reward R_0 associated with the outcome, and a discounted future expected value with the outcome as the new state:

$$\begin{aligned}&\mathrm{Value}(\mathrm{outcome}) = R_0(\mathrm{outcome}) + \\ &\qquad \gamma \cdot max_i[\mathrm{EVC}(\mathrm{signal}_i,\mathrm{outcome})] \quad (4)\end{aligned}$$

The discounted future expected value is defined recursively as the maximum EVC across all feasible control signals with the outcome as the new state times a discount factor γ ranging from 0 to 1, where 1 means there is no discounting, and 0 means the future is fully discounted.

Once the EVC is computed for all feasible control signals for a particular state, the optimal control signal, signal*, is determined by finding the maximum EVC and its corresponding control signal:

$$\mathrm{signal}^* \leftarrow max_i[\mathrm{EVC}(\mathrm{signal}_i,\mathrm{state})] \quad (5)$$

Despite the two elements of the Value function (Eq. 4), previous computational implementations of EVC theory have ignored the discounted future expected value, and only considered immediate reward when computing value (Musslick et al., 2019). This 'greedy' EVC greatly simplifies the EVC calculation, eliminating its recursiveness, and is preferable when there are no predictable changes in the agent's state in the future. However, it fails to account for the value of learning.

LEVC Theory

The LEVC theory determines how control allocation should change when an agent's ability can improve over time. We define an agent's ability in a particular task as its automaticity. Automaticity improves through learning as a function of experience. Some tasks are learnable, and some tasks are not learnable (Fig. 1a). Here, to reduce computational complexity, we have assumed that automaticity on a particular task L increases linearly with experience:

$$\text{automaticity}_L = \alpha_L \cdot n_{\text{trials}\,L} + \text{automaticity}_{0_L} \qquad (6)$$

where automaticity$_{0_L}$is the agent's initial ability at the task, and α_L is its learning rate.

An agent's automaticity in turn determines its accuracy vis-à-vis control signal intensity (Fig. 1b). For example, a skilled pianist may allocate less control for a perfect performance, while an intermediate player will require much more control for a similar performance. Many functions could be used to model this relationship. We use the following sigmoid:

$$\text{accuracy} = \frac{1}{1 + e^{-\text{rate}\cdot(\text{signal}-\text{bias})}} \qquad (7)$$

where rate $=$ automaticity/difficulty of the task, and bias $=$ b_{adj}/rate, where b_{adj} is a free parameter chosen based on the range of control signals used to keep the resulting sigmoids within a comparable range.

To calculate the instantaneous greedy EVC, an accuracy function will be multiplied by the corresponding reward for the current state minus the control cost (Fig. 1c). Different automaticities lead to different EVCs (Fig. 1d).

However, as stated previously (Eq. 6), automaticity can only be increased with experience, which means the agent must choose to perform the task in order to improve on it and later reap the benefits of that improvement. This is where it becomes important to consider the discounted future expected value term in the Value function (Eq. 4). Notably, the value of learning results directly from the discounted future reward obtained from task practice. In this way, the LEVC model attributes the intrinsic value of learning to predicted discounted future reward from learning, without relying on the assumption that learning is inherently valuable in and of itself.

If we compute the greedy EVC on trial 1 for two tasks, one learnable (solid cyan) and one static (solid red), then the static task will yield a greater EVC and the agent will choose to perform that task. If we compute the LEVC for the same tasks, learnable (dotted cyan) and static (dotted red), the agent will instead choose the learnable task (Fig. 1e). Having chosen this task, the agent's automaticity will increase, and on subsequent trials less control will yield equal or greater reward.

LEVC Process Model

Having set up the LEVC optimization problem, we must determine how the agent will make its choices. The recursive nature of the LEVC means that the EVC value of a particular task on a given trial depends both on its previous choices, and on all future possible choices. In this first presentation of the LEVC, we have opted for dense computation of these values for a limited number of trials, and assume that the agent has access to these values when making decisions. Dense computation reveals a normative path under the parameters chosen. Future implementations will incorporate EVC approximations that are less computationally expensive and more biologically plausible.

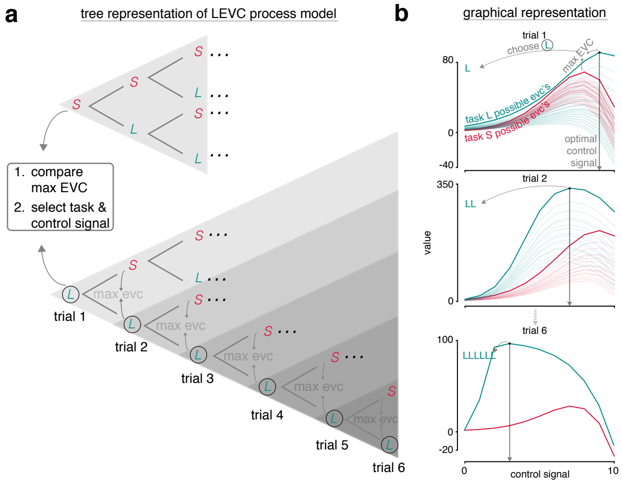

Figure 2: **Learning expected value of control process model. (a)** An agent compares the maximum EVC of all available trajectories on trial 1 of a 6-trial task. Half of these trajectories begin with a static task S, the other half with a learnable task L. The agent selects the task leading the trajectory with the highest maximum EVC and its corresponding optimal control signal. This process is repeated until the end of the task. **(b)** EVC curves for all $2^6 = 64$ possible trajectories on trial 1, colored by whether they begin with task L or S. The curve with the highest EVC for each initial task is indicated with a darker shade. On trial 2, there are now $2^5 = 32$ possible trajectories and the agent picks the task and optimal control intensity corresponding to the one with the highest maximum EVC. By trial 6, only $2^1 = 2$ trajectories are left.

We assume an agent can choose between a learnable task (L), and a static task (S) for a length of n trials. When there are six trials and two choices per trial, there are $2^6 = 64$ possible trajectories on the first trial. Half of these trajectories correspond to an initial choice of task L, and the other half to an initial choice of task S. The agent computes the maximum EVC among the 64 possible trajectories, and chooses the initial task (L) and corresponding optimal control signal of the best trajectory (Fig. 2a-b, trial 1). On the second trial, the agent repeats this process, but because the agent has chosen task L already, it only has $2^5 = 32$ possible trajectories available, all of which include an initial choice of task L (Fig. 2a-b, trial 2). The agent repeats this process until the last trial, when there are only 2 remaining possible trajectories, making a final choice of task L or S. This procedure will culminate in the optimal task and control signal choice trajectory for the given parameters.

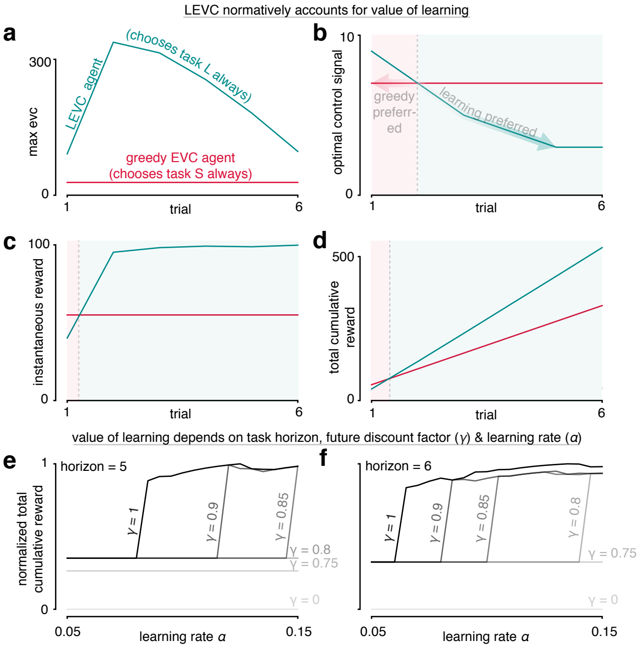

Figure 3: **Control allocation as a function of learnability.** **(a)** Maximum EVC per trial for LEVC (cyan) and greedy (red) EVC agents. **(b)** LEVC agent optimal control signal drops below greedy EVC agent early on. **(c)** & **(d)** LEVC agent starts with lower instantaneous and total cumulative reward, but surpasses greedy EVC agent despite reduced control signal. **(e)** & **(f)** Longer task horizons, higher learning rates α, and larger discount factors γ increase value of learning, resulting in higher total cumulative reward. Total cumulative reward is normalized to compare fairly across horizons (fewer trials means fewer available rewards.)

Results

Control Allocation as a Function of Learnability

To understand the value of learning and its effects on control allocation, we simulated a greedy EVC agent ($\gamma = 0$), and a future discounted LEVC agent ($\gamma = 0.9$) with otherwise equal parameters, meaning learning was available in both cases.

The optimal trajectory for the greedy EVC agent (red) is the choice of the static task S on every trial, whereas the optimal trajectory for the LEVC agent (cyan) is the choice of the learnable task L on every trial. The maximum EVC is higher for all trials for the LEVC agent (Fig. 3a). The non-monotonic shape of the LEVC's maximum EVC curve is a result of the EVC value being partially determined by future discounted rewards. As the agent nears the end of the experiment, there are fewer future rewards to include in its current choices. This observation makes the prediction that if an agent does not anticipate continuing a task, its expected value will decrease. This in turn predicts that learning only has added value when the agent anticipates reaping the value of that learning in the future.

The optimal control signal trajectory starts higher for the LEVC and subsequently drops below the greedy EVC (Fig. 3b). Thus, investing in learning allows an agent to exert less control for equal or greater rewards.

Notably, the LEVC agent chooses the learnable task despite the fact that its initial instantaneous and total cumulative rewards are lower than for the greedy EVC agent (Fig. 3c-d). Considering the entire simulation, choosing the learnable task is a rational choice, as it leads to substantially higher instantaneous rewards, and to a larger total cumulative reward.

The LEVC provides a normative judgment on the value of learning depending on the parameters provided. To probe the value of learning across relevant parameter values, we computed the total cumulative reward for different task time horizons, discount factors and learning rates (Fig. 3e-f). As expected, when the future is fully discounted ($\gamma = 0$), the learning rate and horizon are irrelevant to the agent's task choice, and it cannot reap the benefits of learning. As the future becomes less discounted (higher γ), the agent begins to benefit from the value of learning, *i.e.* the smaller the learning rate needs to be for the agent to choose the learning task. The step-wise changes in total cumulative reward correspond to the agent switching from an optimal trajectory of task S always to one of task L always. An increased horizon exacerbates this effect, leading to even lower learning rates required for the same discount factors. The step-wise changes, indicating a switch in strategy, highlight that an agent's predicted horizon, learning rate and discount factor are crucial in determining its behavior. Notably, these variables can be shaped through suggestion ("95% of participants learned this task") and task design, and their effects tested empirically. Previous work has already demonstrated that subjects will choose information over reward when aware of a longer horizon (Wilson, Geana, White, Ludvig, & Cohen, 2014).

Cognitive Effort Discounting with Learnable Tasks

The COGED task measures the subjective value of a harder task relative to an easier baseline task (Westbrook & Braver, 2015). The subjective value is thus a quantification of the cost of cognitive effort. For example, consider the harder task pays $10 and is 10% more difficult than the baseline task. The experimenter raises the reward of the baseline task until the participant selects the baseline task over the harder task. If the participant switches to the baseline task when it pays $7, then the harder task has a subjective value of $7, or 0.7 relative value units, when it is 10% more difficult.

In order to test predictions of the LEVC theory, we propose and simulate an experiment of cognitive effort discounting with learnable tasks (L-COGED). In the L-COGED task, a harder but learnable (L) task is pitted against an easier but static (S) baseline task. In this situation, a subject's choice and effort directed towards the learnable task has the potential to generate larger future rewards. Such a setup leads to the prediction that the learnable task should have a subjective value above the baseline task, provided the value of learning can be harvested (considering task horizon, discount factor, and learning rate, as seen in Fig. 3e-f).

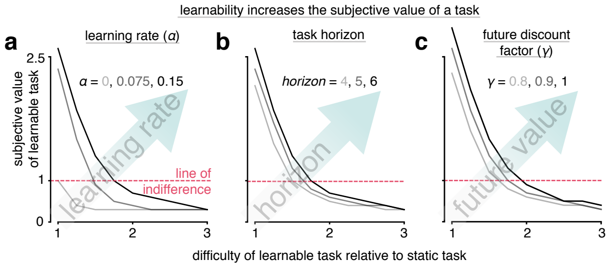

Figure 4: **Cognitive effort discounting with learnable tasks simulation.** To simulate the L-COGED task, we modified two parameters relative to the previous simulations: S task given a lower difficulty than L task (previously equal), and S & L given equal initial automaticities (previously S was higher). **(a)** Subjective value of learnable task across learning rates, **(b)** task horizons, **(c)** and discount factors.

Our simulation of the L-COGED task predicts this behavior. We first replicate the dependency of subjective value on task difficulty, as observed in the regular COGED task with no learning (Fig. 4a, $\alpha = 0$; Westbrook and Braver (2015)). We then show that higher learning rates lead to higher subjective values for the learnable task, above the value of 1 for certain difficulties (Fig. 4a). We observe the same decay in subjective value with increasing difficulty, indicating that as a task becomes more difficult, investing cognitive effort in learning holds increasingly less value.

Accordingly, a longer horizon and larger consideration of future value lead to a higher subjective value in the L-COGED simulation (Fig. 4b-c). A participant's perceived learning rate, discount factor and horizon can be manipulated experimentally through suggestion and task design, allowing us to test these hypotheses. We note, however, that careful experimental design will be required to dissociate these variables as they produce the same effects. Additionally, these simulations may begin to provide an answer for the 'paradox of effort,' the aversiveness of boredom, differences in control allocation related to psychiatric disorders, and the "need for cognition" (Cacioppo & Petty, 1982; Geana, Wilson, Daw, & Cohen, 2016; Grahek, Shenhav, Musslick, Krebs, & Koster, 2019; Inzlicht et al., 2018).

Discussion

We presented a theory that normatively links the value of learning with cognitive control allocation. The LEVC theory shows that it is advisable for an agent to choose to learn, even when that choice requires foregoing present rewards. We further provide predictions for the subjective value of learning that can be tested behaviorally.

Related Work Previous empirical work demonstrated that rats trade present rewards for faster learning, a decision requiring cognitive control (Masís et al., 2020). The authors show this behavior is normative based on a neural network extension of the drift-diffusion model, in which the choice of threshold is an analog of control allocation, but that lacks a process model through which an agent allocates control towards learning. The LEVC builds on the empirical and theoretical predictions in this work and provides a normative framework and process model for how cognitive control is allocated in order to account for learning.

One recent study concerning control allocation and learning dissected the tradeoff between learning efficiency and multiprocessing ability (Sagiv et al., 2020). Assuming an agent will learn, should it use shared or separated representations? Shared representations have been shown to promote learning efficiency. However, once learned, shared representations can cause two tasks to interfere with one another when executed simultaneously (Musslick et al., 2016). Sagiv et al. (2020) showed that, under a wide range of parameter values, ideal agents opt for learning faster using shared task representations, at the expense of multitasking capability. Our study instead focuses on the question of whether an agent should choose to learn. Future extensions of the LEVC could allow agents to split control among tasks. Such a framework would allow a richer exploration of the bilateral mechanisms through which learning generates value: improvement in automaticity (explored here), and the concomitant reduction in interference, leading to a reduced cost of control. This framework would also allow examining questions of training regimes, such as interleaved or blocked learning, which have been shown to affect learning efficiency (Flesch, Balaguer, Dekker, Nili, & Summerfield, 2018).

One extension of EVC, the learned-value-of-control (LVOC) theory, specifies a method through which an agent can approximate the optimal allocation of control for a particular task environment given previous experience in similar task environments (Lieder, Shenhav, Musslick, & Griffiths, 2018). This method addresses the formidable computational complexity of calculating an optimal control signal on the fly. The LEVC Theory presented here, by contrast, focuses on allocating optimal cognitive control when an agent can improve at a task through learning. Future work could combine both extensions of EVC theory (Shenhav et al., 2013), solving the problem of calculating optimal control on the fly while ensuring that that control reflects learning prospects in the future.

A recent elaboration of the EVC theory posits that people take into account the efficacy of their cognitive effort when choosing how to allocate control (Musslick, Cohen, & Shenhav, 2018), and was recently empirically examined (Frömer, Lin, Dean Wolf, Inzlicht, & Shenhav, 2020). For example, a game of blackjack with perfect strategy only yields 49:51 odds, so it may not be worth the control required to play the game perfectly. The role of control efficacy in the EVC is closely related and complementary to our learning extension, as learning can be conceptualized as a strategy an agent can pursue in order to increase its control efficacy.

Intrinsic Value of Learning Whether learning is intrinsically valuable to biological agents is an open question. Neural data supports the idea that agents may treat information (a consequence of learning) as a good in itself: a heuristic for its future discounted reward over the lifetime of the agent. (Bromberg-Martin & Hikosaka, 2009; Kang et al., 2009; Gottlieb & Oudeyer, 2018). In line with this view, work in reinforcement learning posits that such a value of information is proportional to how that information can predict future rewards (Behrens, Woolrich, Walton, & Rushworth, 2007). The LEVC model presented here, however, propounds that the value of learning can be specified entirely through predicted discounted future rewards resulting from that learning, without the need for a separate parameter encoding that value. Because an LEVC agent has access to how its actions might change its future self, there is no need for an intermediary "value of information" term. That value is taken into account when estimating future discounted rewards. This framing generates the prediction that learning has added value if it can be applied in the future (3a). Nonetheless, one could argue that the intrinsic value of learning (and effort, Inzlicht et al. (2018)) are in fact directly available to humans: the intractability of computing discounted future rewards may require the system to cache the associated value of learning into some "intrinsic value". Such a cached value for learning could be conceived as a prior representing the discounted future rewards from previous learning experiences.

Learnability One outstanding question is how an agent ascertains a task's learnability. One possible method is for the agent to estimate learnability as a function of how predictable it finds its environment. In curiosity-driven reinforcement learning, an agent is rewarded for how poorly it predicts its environment, pushing it towards constant exploration (Burda et al., 2019). Novelty, however, is not equal to learnability: a maximally random environment, such as a static-filled TV, will instigate curiosity but it is inherently unlearnable. Another way an agent could estimate task learnability is through experience, extrapolating improvements with task practice into the future (Ravi et al., 2020). Yet another related way an agent could estimate task learnability is by comparing its learning rate against an optimal learning rate, analytically available in some cases (Wilson, Shenhav, Straccia, & Cohen, 2019), and conceivably estimated based on experience. Such a method could inform the agent not only if a task is learnable, but if it is worthwhile. For instance, some tasks, such as those requiring data superseding an agent's processing capacity, remain unlearnable, which would lead to learning rates far from optimal. This growing body of work suggests that learnability is knowable, or at least that it can be estimated. In LEVC, we assume that the agent already has an estimate of the task learnability, and the question we seek to answer is how it allocates control once it does. Future work will focus on allowing the LEVC agent to develop that estimate of learnability based on its interactions with the environment.

Clinical Implications The value of learning may contribute to the understanding of cognitive impairments in psychiatric disorders, as commonly observed in control-demanding scenarios (Grahek et al., 2019). A vast amount of research in computational psychiatry finds that psychiatric dysfunctions are associated with perturbations in reinforcement learning, such as schizophrenia patients showing selective impairments in reward-driven learning (Waltz, Frank, Robinson, & Gold, 2007). While perturbations in learning behavior are amenable to computational analysis, other psychiatric impairments, such as ones associated with deficits in cognitive control, remain less well understood (Millan et al., 2012). EVC theory offers a possible explanation for the cognitive deficits of depressive patients in control-demanding tasks, suggesting that depression may be associated with a higher cost of cognitive control (Grahek et al., 2019). The present study suggests that these impairments may as well result from a reduced value of learning, linking psychiatric perturbations in learning with perturbations in effort allocation.

Future Work The learning algorithm described in this article relies on the assumption that the performance of a cognitive agent improves as long as they engage with the task, irrespective of their actual performance. We adopted this assumption to reduce the computational complexity of computing EVC: future performance is only dependent on the selected task (control identity), not performance (control intensity). However, in realistic settings, biological and simulated agents are likely to learn as a function of how well they perform on a task (Masís et al., 2020). Thus, future implementations of LEVC should explore performance-dependent learning mechanisms. The resulting increase in computational complexity may require approximating EVC, by, for instance, learning the optimal control policy based on reinforcement. This extension would allow, for instance, understanding from a control point of view the prediction that participants should seek optimal learning rates, and should otherwise prefer alternative tasks (Wilson et al., 2019).

Acknowledgments

This project was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. J.M. was supported by the Presidential Postdoctoral Research Fellowship at Princeton University. We thank Andrew Saxe for helpful discussions.

References

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature neuroscience*, *10*(9), 1214–1221.

Bromberg-Martin, E. S., & Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*, *63*(1), 119–126.

Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T., & Efros, A. A. (2019). Large-scale study of curiosity-driven learning. In *International Conference on Learning Representations (ICLR).*

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of personality and social psychology*, *42*(1), 116.

Flesch, T., Balaguer, J., Dekker, R., Nili, H., & Summerfield, C. (2018). Comparing continual task learning in minds and machines. *Proceedings of the National Academy of Sciences*, *115*(44), E10313–E10322.

Frömer, R., Lin, H., Dean Wolf, C. K., Inzlicht, M., & Shenhav, A. (2020). When effort matters: Expectations of reward and efficacy guide cognitive control allocation. *bioRxiv*. doi: 10.1101/2020.05.14.095935

Geana, A., Wilson, R., Daw, N. D., & Cohen, J. D. (2016). Boredom, information-seeking and exploration. *In Proceedings of the 38th Annual Meeting of the Cognitive Science Society. Cognitive Science Society*, 1751-1756.

Gottlieb, J., & Oudeyer, P.-Y. (2018). Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience*, *19*(12), 758–770.

Grahek, I., Shenhav, A., Musslick, S., Krebs, R. M., & Koster, E. H. (2019). Motivation and cognitive control in depression. *Neuroscience & Biobehavioral Reviews*, *102*, 371–381.

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in cognitive sciences*, *22*(4), 337–349.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological science*, *20*(8), 963–973.

Kool, W., & Botvinick, M. (2018). Mental labour. *Nature human behaviour*, *2*(12), 899–908.

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. *The Behavioral and brain sciences*, *36*(6).

Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. *PLoS computational biology*, *14*(4), e1006043.

Masís, J., Chapman, T., Rhee, J. Y., Cox, D. D., & Saxe, A. M. (2020). Rats strategically manage learning during perceptual decision making. *bioRxiv*. doi: 10.1101/2020.09.01.259911

Millan, M. J., Agid, Y., Brüne, M., Bullmore, E. T., Carter, C. S., Clayton, N. S., ... others (2012). Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nature reviews Drug discovery*, *11*(2), 141–168.

Musslick, S., Cohen, J. D., & Shenhav, A. (2018). Estimating the costs of cognitive control from task performance: theoretical validation and potential pitfalls. In *Proceedings of the 40th annual conference of the Cognitive Science Society* (pp. 800–805). Madison, WI.

Musslick, S., Cohen, J. D., & Shenhav, A. (2019). Decomposing individual differences in cognitive control: a model-based approach. *In Proceedings of the 41st Annual Meeting of the Cognitive Science Society. Cognitive Science Society, Montreal, CA*, 2427-2433.

Musslick, S., Dey, B., Özcimder, K., Patwary, M., Willke, T. L., & Cohen, J. D. (2016). Controlled vs. automatic processing: A graph-theoretic approach to the analysis of serial vs. parallel processing in neural network architectures. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 1547—1552). Philadelphia, PA.

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.*

Ravi, S., Musslick, S., Hamin, M., Willke, T. L., & Cohen, J. D. (2020). Navigating the trade-off between multi-task learning and learning to multitask in deep neural networks. *arXiv preprint arXiv:2007.10527*.

Sagiv, Y., Musslick, S., Niv, Y., & Cohen, J. D. (2020). Efficiency of learning vs. processing: Towards a normative theory of multitasking. *arXiv preprint arXiv:2007.03124*.

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.

Waltz, J. A., Frank, M. J., Robinson, B. M., & Gold, J. M. (2007). Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. *Biological psychiatry*, *62*(7), 756–764.

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 395–415.

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.

Wilson, R. C., Shenhav, A., Straccia, M., & Cohen, J. D. (2019). The eighty five percent rule for optimal learning. *Nature communications*, *10*(1), 1–9.