ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/312125287

A Computational Model of Control Allocation based on the Expected Value of Control

Conference Paper · June 2015

CITATIONS
25

READS
1,073

4 authors, including:

Sebastian Musslick
Princeton University

43 PUBLICATIONS **487** CITATIONS

SEE PROFILE

Amitai Shenhav
Brown University

80 PUBLICATIONS **3,289** CITATIONS

SEE PROFILE

Matthew M Botvinick
Princeton University

99 PUBLICATIONS **27,173** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Emotional Response Inhibition and the functional theory of the p factor: A computational neuroscience approach to transdiagnostic risk for psychopathology View project

Neural network modeling View project

All content following this page was uploaded by Sebastian Musslick on 07 January 2017.

The user has requested enhancement of the downloaded file.

A computational model of control allocation based on the Expected Value of Control

Sebastian Musslick
Princeton Neuroscience Institute
Princeton University
Princeton, NJ 08540
musslick@princeton.edu

Amitai Shenhav
Princeton Neuroscience Institute
Princeton University
Princeton, NJ 08540
ashenhav@princeton.edu

Matthew M. Botvinick
Princeton Neuroscience Institute
Princeton University
Princeton, NJ 08540
matthewb@princeton.edu

Jonathan D. Cohen
Princeton Neuroscience Institute
Princeton University
Princeton, NJ 08540
jdc@princeton.edu

Abstract

While cognitive control has long been known to adjust flexibly in response to signals like errors or conflict, when and how the decision is made to adjust control remains an open question. Recently, Shenhav and colleagues (*1*) described a theoretical framework whereby control allocation follows from a reward optimization process, according to which the identities and intensities of potential cognitive control signals are selected so as to maximize expected reward, while at the same time discounting this reward by an intrinsic cost that attaches to increases in control allocation. This discounted expected reward quantity is referred to as the Expected Value of Control (EVC). While the form of the reward optimization policy was described, Shenhav *et al.* left open the question of how this optimization process might be implemented in explicit computational mechanisms, and used to make predictions concerning performance in experimental tasks. Here we describe such an implementation.

To simulate the influence of cognitive control on behavior in relevant task settings we parameterize such tasks as processes of accumulation to bound and allow control to influence the parameters of that accumulation process, resulting in attendant changes in reward rate. Control signals are specified based on an internal model of the task environment, as well as the intrinsic cost of control allocation. The latter scales both with the amount of overall control exerted and with the change in control allocation from the previous time step. After control is applied, feedback from the actual task environment is used to update the internal model, and control specification is re-optimized. We show that the behavior of a simulated agent using these mechanisms replicates classic findings in the cognitive control literature related to sequential adaptation and task switching, and is able to generate testable predictions for future studies of voluntary rather than instructed allocation of control.

Keywords: cognitive control, sequential adaptation, task switching, drift-diffusion model

1 Introduction

An essential component of human cognitive flexibility is our ability to override habitual responses in order to successfully control behavior in the service of current task goals. Mechanisms underlying this ability can be summarized under the term *cognitive control*. Early computational work proposed that control units serve to strengthen task-relevant processing pathways in order to reduce interference from competing responses (*2*). Subsequent work suggested that the intensity of a particular control signal might be adjusted in response to the detection of conflict in processing, as a signal of the need for control (*3*). This suggested that, in addition to its regulatory function, cognitive control would also require a monitoring component that provides feedback when control needs to be recruited.

In order to achieve optimal performance, the implementation of control requires a preceding decision about how to allocate control. This must include specification of both the identity of a candidate control signal (e.g., selection of a particular task rule) and its intensity (that is the strength of the signal, or *how much* control should be allocated). Shenhav *et al.* provided a normative account for the specification of such properties of control signals by integrating over the expected payoffs (e.g., monetary reward) as well as the intrinsic costs individuals associate with the exertion of cognitive effort (which is assumed to scale with both the intensity of the control signal allocated and its distance from the previously allocated signal). The resulting integrated quantity, referred to as the Expected Value of Control (EVC), is theorized to serve as the basis for decisions about control allocation: Control is allocated to tasks with the highest EVC.

While Shenhav and colleagues offered a formal description for specifying control signals that maximize EVC, they left open the question of how this process of specification might be implemented in computational mechanisms. Here we introduce a computational framework for implementing EVC-based control specification in actual cognitive tasks. We show how such control adjustments can explain classic cognitive control phenomena, including those related to sequential adaptation (in response to errors or task conflict) and task switching, as well as more recent findings related to the influence of reward on task switching. We also describe testable predictions the model makes within these domains, which are the target of ongoing empirical investigation.

2 Methods

In simulations using the EVC model, each trial can be described as an interaction between the control system and the task environment. The system specifies the optimal control signal based on an internal representation that is an estimate of the task environment (inferred state $\hat{S}$). After implementing the signal and performing the corresponding task(s) in the actual task environment (actual state S), the model receives feedback based on its performance, which is then used to update the inferred state for the next trial.

2.1 Parameterization of the Task Environment and Control Signals

We parameterize the initial state S for each trial in terms of a set of sensory stimuli with each stimulus being associated with one of two potential responses that lead to corresponding outcome states $O_i, i \in \{1, 2\}$. A simulation of the Erikson Flanker Task (*4*,*5*) would involve a set of arrow stimuli (pointing either to the left or right) that drive an internal process of accumulating evidence to bound for the two potential outcome states (after correct or incorrect response), one of which may be strongly (automatically) biased as a result of previous learning (e.g., responding to the majority of arrows on the screen rather than a central one as instructed). We simulate this process with the Drift-Diffusion Model (DDM) (*6*). In order to perform the task optimally, a control signal u would have to act on the central target stimulus by driving its associated response. Control can be parameterized either as a modification of the drift rate or threshold parameter whereas the (automatic) bias (b) represents a component of the drift rate. Finally each of the outcome states (after correct or incorrect response) is associated with a value $V(O_i)$.

2.2 Control Signal Specification

For a given control configuration $\overrightarrow{u}$ in a multi-dimensional signal space the expected value for the current trial t is determined as a function of the expected reward rate (given the inferred state $\hat{S}$) weighted against a cost term[1]:

$$EVC(\overrightarrow{u}, \hat{S}_t) = \left[\sum_{i=1}^{2} P(O_i | \overrightarrow{u}, \hat{S}_t) \frac{\hat{V}(O_i)_t}{1 + RT(O_i | \overrightarrow{u}, \hat{S}_t)} \right] - Cost(\overrightarrow{u}) \tag{1}$$

[1] Note that the formula in Eq. 1 is a special case of the general EVC formula as proposed in (*1*) in that our implementation represents a 1-step look-ahead policy rather than a recursive policy discounting future rewards. However, future simulations will seek to explore discounting over longer horizons.

For both outcome states O_i we can use the DDM parameterization to compute the probability $P(O_i|\overrightarrow{u}, \hat{S}_t)$ of reaching the state, as well as the required reaction time $RT(O_i|\overrightarrow{u}, \hat{S}_t)$ in order to obtain the expected reward rate[2]. The Cost term is composed of an implementation cost $f(x)$ and a reconfiguration cost $g(x)$:

$$Cost(\overrightarrow{u}_t) = f(\overrightarrow{u}_t) + g\left(\sqrt{\sum_j^n (u_{j,t} - u_{j,t-1})^2}\right) \tag{2}$$

where $f(x)$ and $g(x)$ are both exponential functions. Note that while the implementation cost depends on the *absolute* signal intensity (i.e., how much total control is being allocated), the configuration cost is context dependent — that is, it depends on the previous control signal. Specifically, it is computed as the Euclidian distance in signal space between the candidate control configuration at the current trial t and the previously implemented control signal configuration at trial $t-1$ (i.e., how much control would need to be adjusted). This reconfiguration term is included to account for potential costs of switching signals. Among all possible configurations, the model chooses the control signal configuration with the maximum EVC:

$$\overrightarrow{u}_t^* \leftarrow argmax_i[EVC(\overrightarrow{u_i}, \hat{S}_t)] \tag{3}$$

2.3 Control Signal Implementation and Update of the Expected State

After the optimal control signal is specified based on the inferred state of the environment, this signal is implemented in a DDM (e.g., by adjusting the drift rate) that influences the actual state, and the agent receives feedback about the resulting outcome probabilities[3] $P(O_i|\overrightarrow{u^*}, S_t)$. The parameter representing the inferred automatic bias is then updated according to the following rule:

$$\hat{b}_{t+1} = \hat{b}_t - \alpha_b(P(O_1|\overrightarrow{u^*}, S_t) - P(O_1|\overrightarrow{u^*}, \hat{S}_t)) \tag{4}$$

where α_b is the learning rate for the inferred automatic bias $\hat{b}$ and O_1 is the outcome associated with reaching the lower DDM threshold (representing e.g. an incorrect response). The inferred outcome value representations $\hat{V}(O_i)$ are updated as follows:

$$\hat{V}(O_i)_{t+1} = \hat{V}(O_i)_t + \alpha_v(P(O_i|\overrightarrow{u^*}, S_t)(V(O_i)_t - \hat{V}(O_i)_t) \tag{5}$$

where α_v is the learning rate for the inferred outcome value $\hat{V}(O_i)$.

3 Results

3.1 Specification of control signal intensity

Similarly to previous models of sequential adaptation of control (*3*), the EVC model can account for how individuals dynamically adjust control from trial to trial in order to optimize reward rate (see Fig. 1A-C). For instance, by optimizing over potential specifications of response threshold (caution) the model can reproduce the adjustment that has been observed following the commission of an error (*7*), resulting in longer response times (RTs) and lower error likelihood on the subsequent trial. The model also replicates control adjustments that occur in response to task conflict (*8*). After an incongruent trial the control system chooses to implement a higher control signal intensity (in this case associated with increased drift rate toward the controlled response) leading to faster RTs and fewer errors. Apart from adjustments related to errors and conflict, the model produces adjustments in control based on changes in incentives (*9*). Thus, higher incentives (e.g., reward for a correct response) will tend to elicit increased control allocation, though this effect will depend critically on the costs associated with increasing control.

[2]For current simulations the obtained outcome value gets fully discounted by the RT. However, our model also allows for partial RT discounting.

[3]Probability-based trial feedback was implemented for ease of analysis. The same results obtain when using binary feedback (correct or incorrect).

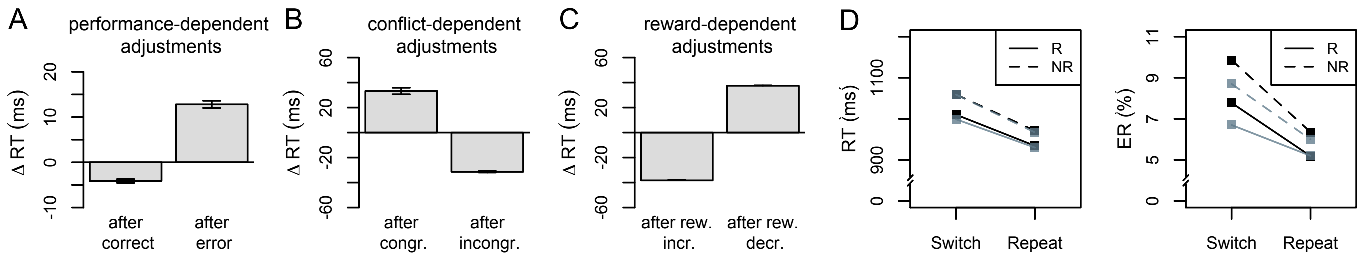

Figure 1: Simulated control adaptation effects and task switching effects. Figures (**A**), (**B**) and (**C**) show trial-by-trial control adaptation effects measured as the mean difference between RTs at trial t and RTs at trial $t-1$. RT adjustments depended on (A) the occurrence of errors in the previous trial, (B) congruency of target and distractor stimuli in the previous trial and (C) unexpected increases or decreases in outcome values for correct responses in the previous trial. The error bars indicate the standard error across simulations with different task sequences. Figure (**D**) shows simulation results of a task switching experiment in which the agent had to switch between a rewarded (R) and non-rewarded task (NR). Mean simulated RTs and ERs (black lines) for the interaction of transition (repetition/switch) and task (rewarded/non-rewarded) are contrasted with empirical data (superimposed grey lines) from (*10*).

3.2 Specification of control signal identity

In task switching experiments participants are instructed to switch between two or more tasks over the course of an experimental session (e.g., judge a number based on its magnitude [greater/less than 5] or parity [odd/even]). The most consistent finding from this literature is that subjects are faster and more accurate at performing a task that is the same as on the previous trial (repeat) versus one that differs from the previous trial (switch) (*11–13*). In other words, switch trials incur a performance cost relative to repeat trials. We simulate such a task environment — by cueing reversal of the reward structure such that the model could anticipate the currently relevant (rewarded) task — and demonstrate robust switch costs for both RTs and error rate (ER) when alternating between the control signals relevant to the two tasks.

A recent study by Umemoto *et al.* (*10*) investigated the influence of reward on task performance and task switching, in the latter case comparing the costs of switching from a reinforced to a non-reinforced task and vice versa (Fig. 1D). The EVC model captures the results of the study: (a) the reinforced task entails faster reaction times and fewer errors than the non-reinforced task and (b) switch costs were smaller for the non-reinforced task relative to the reinforced task (interestingly, both the model and empirical data show these reduced switch costs occurring only for ERs and not RTs).

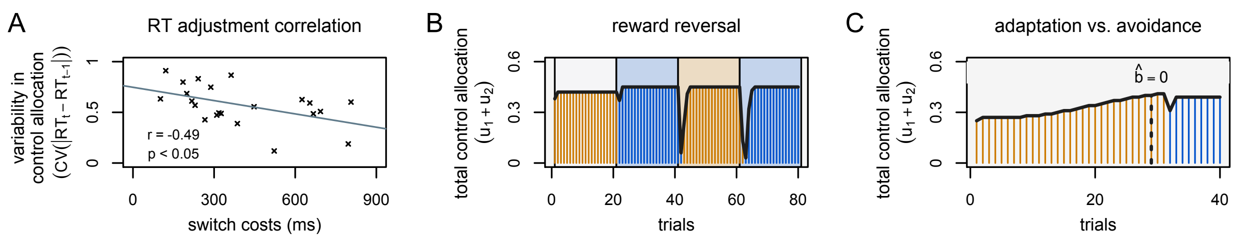

Figure 2: Model predictions for future task switching experiments. (**A**) The coefficient of variance in absolute RT differences from trial to trial within a task block is plotted against RT switch costs that occur between task blocks for different simulations, cf. (*14*). (**B**) The stacked bar plot shows the total amount of control intensity $sum(\overrightarrow{u})$ applied for every trial in a voluntary task switching (VTS) simulation. The control signal space encompassed two signals, one for an easier task (u_1, orange) and one for a more difficult task (u_2, blue). The corresponding background colors indicate the currently more rewarded task while the grey background for the first 20 trials indicates equal outcome values for both tasks. The amount with which the overall control intensity $sum(u)$ is composed of u_1 and u_2 reflects the model's choice of the corresponding task. (**C**) The plot shows a VTS simulation similar to (B). Both tasks were equally rewarded across all trials. However, the automatic bias b indexing the relative difficulty of the two tasks gradually changed from facilitating the orange task towards facilitating the blue task. The dashed line marks the trial at which the inferred bias $\hat{b}$ switched towards the blue task. Continuing adaptation of the orange signal past this point indicates an effect of reconfiguration costs.

In addition to the task switching effects reported above, the model yielded a negative correlation between control adjustment effects within a series of trials of the same required task and switch costs that occurred when the system had to switch to an alternative task (see Fig. 2A). That is, in a blocked task switching paradigm larger switch costs between task blocks are associated with less variability in adaptation effects such as post-error slowing within a single task block for both RTs and ERs. The magnitude of both effects depended on the parameterization of the reconfiguration cost function $g(x)$ which is a function of the relative distance between the previously implemented and the new control signal configuration. This correlation could be examined experimentally to validate the model's assumptions (and/or suggest potential revisions), as well as to constrain parameter values for a given participant.

Whereas most task switching paradigms explicitly cue participants to switch between tasks in a predetermined order, the EVC model can be used to address the wide range of real-world settings in which the agent is able to choose which tasks to perform, how long to perform them, and with what amount of control (see Fig. 2B & 2C). For instance, in a two-task environment, simulation results reveal that the EVC model consistently decided to perform the task that requires less control intensity in order to obtain the same amount of reward. However, sufficiently increasing reward for performing the more difficult task induced the model to switch to this previously avoided task (Fig. 2B). The model also exhibited a non-monotonic response to task difficulty (Fig. 2C), reflecting a natural tension between conflict adaptation and conflict avoidance (*15*): When the difficulty of the preferred task was increased the model initially chose to increase control on that task in order to improve performance. However, if difficulty continued to increase, the model eventually chose to switch to the other task. These behaviors reflect the model's ability to flexibly account for a number of potential dynamic influences on control allocation in real-world decisions about voluntary control allocation, providing further avenues for testing model predictions and empirically constraining its parameters (as discussed below).

4 Discussion

We introduced a computational model of control signal specification based on the Expected Value of Control theory (*1*). The model not only accounts for a variety of classic phenomena, including observations of conflict adaptation and task switch costs, but also yields new predictions for task switching experiments that can be empirically tested. Ongoing work seeks to leverage the task switching paradigm in order to fit individual-level parameter values (based on performance in a cued/instructed task switching environment) and then test predictions concerning the allocation of control and performance of those individuals in a voluntary task switching environment (involving variations in task payoff and difficulty). Subsequent work will seek to further characterize/quantify longer-term dynamics of voluntary control allocation (e.g., cognitive fatigue or "depletion"-like effects). Future work will also seek to constrain model parameters with experimental data that reflect the form of relevant reward and cost functions, and to compare the predictions of this model to alternative models in which control is adjusted based on fewer evaluative inputs (e.g., pure conflict monitoring). A further goal is to implement the model using a more biologically realistic architecture (e.g., neural network) that can be used to further test the model using neural data.

In summary, the current implementation of the EVC theory validates its utility as a model of the evaluation and adjustment of cognitive control allocation. This model builds on previous models of cognitive control by implanting an evaluative/monitoring system that considers not only the demands and/or rewards for control, but also the intrinsic costs associated with the allocation of control. Furthermore, the model allows such evaluations to guide both the types and amounts of control that are applied concurrently. Future work, including ongoing experiments described above, will seek to test and further refine this implementation of the EVC theory.

References

1. A. Shenhav, M. M. Botvinick, J. D. Cohen, *Neuron* **79**, 217 (2013).
2. J. D. Cohen, K. Dunbar, J. L. McClelland, *Psychological review* **97**, 332 (1990).
3. M. M. Botvinick, T. S. Braver, D. M. Barch, C. S. Carter, J. D. Cohen, *Psychological review* **108**, 624 (2001).
4. B. A. Eriksen, C. W. Eriksen, *Perception & psychophysics* **16**, 143 (1974).
5. B. Kopp, U. Mattler, F. Rist, *Psychiatry research* **53**, 129 (1994).
6. R. Ratcliff, *Psychological review* **85**, 59 (1978).
7. D. R. J. Laming, *Information theory of choice-reaction times* (Academic Press, London, 1968).
8. G. Gratton, M. G. H. Coles, E. Donchin, *Journal of Experimental Psychology: General* **121**, 480 (1992).
9. S. Padmala, L. Pessoa, *Neuropsychologia* **48**, 558 (2010).
10. A. Umemoto, C. B. Holroyd, *Psychological Research* pp. 1–10 (2014).
11. A. Allport, E. A. Styles, S. Hsieh, *Attention and Performance XV*, C. Umiltà, M. Moscovitch, eds. (The MIT Press, Cambridge, MA, 1994), pp. 421–452.
12. A. T. Jersild, *Archives of Psychology, Whole No. 89* (1927).
13. R. D. Rogers, S. Monsell, *Journal of Experimental Psychology: General* **124**, 207 (1995).
14. M. Esterman, S. Noonan, M. Rosenberg, *Journal of Vision* **12**, 383 (2012).
15. M. M. Botvinick, *Cognitive, affective & behavioral neuroscience* **7**, 356 (2007).

View publication stats