## Title
Estimating the costs of cognitive control from task performance: theoretical validation and potential pitfalls



**Journal**
Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

**Authors**
Musslick, Sebastian
Cohen, Jonathan D
Shenhav, Amitai

**Publication Date**
2018

# Estimating the costs of cognitive control from task performance: theoretical validation and potential pitfalls

Sebastian Musslick[1,*], Jonathan D. Cohen[1,2], and Amitai Shenhav[3]

[1]Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA.
[2]Department of Psychology, Princeton University, Princeton, NJ 08544, USA.
[3]Department of Cognitive, Linguistic, and Psychological Sciences,
Brown Institute for Brain Science, Brown University, Providence, RI 02912, USA.
[*]Corresponding Author: *musslick@princeton.edu*

## Abstract

Cognitive control is critical for accomplishing daily tasks and yet we experience it as effortful or costly. Researchers have been increasingly interested in estimating how costly cognitive control is for a given individual, to better understand underlying mechanisms and predict motivational impairments outside the lab. Here we leverage a computational model of control allocation to (a) demonstrate a procedure for estimating individual's control costs from task performance and (b) highlight the conditions under which estimated costs will be confounded with other motivational variables. We show that costs of cognitive control can be reliably estimated under perfect assumptions about other motivational variables. However, our simulation results indicate that poorly calibrated estimates of those other variables can lead to potentially drastic misestimations of subjects' control costs, compromising the validity of empirical observations. We conclude by discussing the implications of these analyses for assessing individual differences in the costs of cognitive control.

**Keywords:** mental effort; individual differences; cognitive control; expected value of control

## Introduction

Everyday we are confronted with tasks that require us to flexibly bias the processing of stimuli in accordance with relevant task goals, engaging mechanisms that are referred to as cognitive control. Despite its benefits, people seem to avoid exerting cognitive control, suggesting that it is associated with a cost (Botvinick & Braver, 2015; Shenhav et al., 2017).

The general observation that participants can increase the amount of cognitive control allocated to a task if worth the incentive, but generally hold back from doing so, has lead to the assumption that cognitive control is associated with an intrinsic subjective cost (Botvinick & Braver, 2015). For instance, participants respond faster and more accurately on a cognitive control task (e.g. name the ink of a color word instead of reading the word) when offered a greater reward for their performance (Krebs, Boehler, & Woldorff, 2010). Similarly, task switching performance (Umemoto & Holroyd, 2015) and selective attention (Padmala & Pessoa, 2011) seem to improve if higher incentives are offered. The costs of control are also reflected in participant's preferences for tasks. Individuals choose to avoid task switching sequences with a higher demand for task switches (Kool, McGuire, Rosen, & Botvinick, 2010) and are willing to forgo rewards to avoid tasks that impose a higher working memory load (Westbrook, Kester, & Braver, 2013). These and other findings have led to the development of theories that suggest that control allocation follows from a cost-benefit analysis (Shenhav, Botvinick, & Cohen, 2013; Kurzban, Duckworth, Kable, & Myers, 2013).

Researchers have grown increasing interest in measuring the cost of control across individuals in order to predict behavior outside of the lab. Proxies of control costs, such as demand avoidance, have been reported to correlate with IQ (Gold et al., 2015), need for cognition (Westbrook & Braver, 2015), as well as measures of self control (Kool, McGuire, Wang, & Botvinick, 2013) and negative symptom severity in schizophrenia (Barch, Treadway, & Schoen, 2014). However, indices for the cost of control that are estimated from subjects' behavior could instead reflect measures of other, confounded motivational variables (e.g. the subject's sensitivity to reward). The question arises: Does higher demand avoidance or reduced task performance for a given subject reflect a higher cost of control, a lower motivation or a reduced capability to perform the task?

In this work we utilize the expected value of control (EVC) model (Shenhav et al., 2013; Musslick, Shenhav, Botvinick, & Cohen, 2015) as an economically informed theory of control allocation, to derive a method for estimating control costs based on subject's task performance. Validating this method, we show that an individual's control cost function can be estimated from task performance under correct assumptions about other motivational variables. However, we also use the model to expose how incorrect assumptions about these variables can lead to misestimations of control costs, limiting their predictive validity with respect to out-of-lab performance. The code for all computational simulations can be accessed at `github.com/musslick/CogSci-2018a`.

## Control Cost Estimation Based on Expected Value of Control Theory

The EVC theory by Shenhav et al. (2013) proposes that the optimal amount of control is determined by maximizing the expected value of control, that is, the expected utility of implementing a control signal with a given intensity $u$ minus an intrinsic cost that scales with the intensity of the signal

$$EVC(u,S) = \sum_{i=1}^{n} \left(P(O_i|u,S)V(O_i)\right) - Cost(u) \qquad (1)$$

where $P(O_i|u,S)$ is the probability of achieving outcome $O_i$ (e.g. correct response) with $i \in (1,...,n)$ where $n$ is the number of possible outcomes, $V(O_i)$ is the perceived value of reaching that outcome (e.g. the monetary value of responding correctly) and $Cost(u)$ is the cost of implementing the control signal with intensity $u$. It is hypothesized that the control system chooses to implement the control signal with the maximal expected value of control:

$$u^* = \underset{u}{\operatorname{argmax}} EVC(u,S) \quad (2)$$

Provided that participants' behavior is a result of the reward maximization process formulated in Equations 1 and 2, it is possible to solve for the derivative of $Cost(u)$

$$0 = \frac{dEVC}{du} = \sum_{i=1}^{n} \left( \frac{dP(O_i|u,S)V(O_i)}{du} \right) - \frac{dCost(u)}{du} \quad (3)$$

$$\frac{Cost(u)}{du} = \sum_{i=1}^{n} \left( \frac{dP(O_i|u,S)}{du} V(O_i) \right). \quad (4)$$

In experimental settings where there is only one correct (and rewarded) response to a given stimulus, i.e. $V(O_{correct}) > 0$ and $V(O_i) = 0$ with $i \in (2,...,n)$, Equation 4 reduces to

$$\frac{dCost(u)}{du} = \frac{dP(O_{correct}|u,S)}{du} V(O_{correct}). \quad (5)$$

With this setup it is possible to compute the first order derivative of the cost of control function $dCost(u)/du$ from the value of the reward provided for responding correctly $V(O_{correct})$, as well as the derivative of participants' accuracy $P(O_{correct}|u,S)$. The functional form of the cost function can be approximated by integrating $dCost(u)/du$.

Estimating the functional form of $Cost(u)$ requires sampling multiple values of $dP(O_{correct}|u,S)/du$. This can be achieved experimentally by manipulating the amount of reward an agent receives for responding correctly on the same task. Each reward condition $k$ is associated with an optimal control signal $u_k^*$ and a corresponding task accuracy $P(O_{correct}|u_k^*,S)$, leading to different samples of $Cost(u_k)$. Note that this procedure can be applied to any experimental setting with a fixed trial difficulty and varying reward structure where, in its simplest form, there is only one correct, rewarded response associated with each trial (e.g., in the Stroop task only the response indicating the ink color of the word is considered correct and rewarded). These criteria can be met by most paradigms that are used to assess controlled behavior.

The control cost estimation procedure based on the EVC model exposes what information an experimenter would need to estimate control costs. This includes the subject's task accuracy as a function of her control signal intensity $P(O_{correct}|u,S)$, as well as her subjective value as a function of reward $V(O_{correct})$. The former is a function of task automaticity, i.e. the ability to perform the task via automatic processes, without allocation of control. The latter depends on the subject's sensitivity to monetary incentives (reward sensitivity), as well as on the perceived value of performing well on the task without external incentives (accuracy bias). Attempting to make inferences about control costs without incorporating these factors can be problematic. In the next section we will define a ground truth for these variables before going onto exploring how sensitive control cost estimates are to measurement error in these variables.

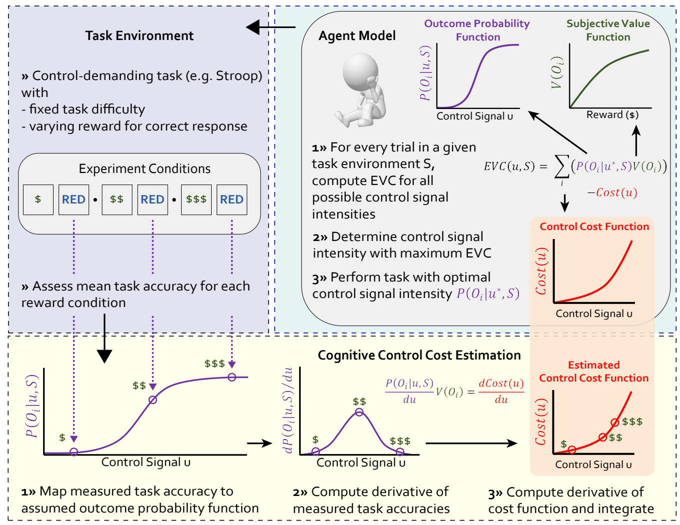

Figure 1: Control cost estimation procedure. The goal is to infer the agent's true subjective cost that scales with the amount of control $u$ allocated to a task (shown in red). Estimating these costs involves measuring the agent's performance on a given task across different reward conditions and computing the derivative of the agent's cost of control for each reward condition.

## Parameterization of Agents and Task Environment

An EVC agent is assumed to allocate control by taking into account an intrinsic cost that scales with control signal intensity. While Shenhav et al. (2013) do not commit to any particular functional form, they suggest that the cost of control increases monotonically with the amount of control allocated to a task. In the simulations below we chose an exponential cost function as the ground truth for a given agent $j$ as

$$Cost(u) = e^{c_j u} - 1 \quad (6)$$

where the cost parameter $c_j$ scales the increase in cost of control with one unit of control signal intensity $u$. Note that the parameterization of $Cost(u)$ may vary across different types of control signals and can, in principle, differ across cognitive control paradigms. We will also validate the estimation of other functional forms, such as quadratic $Cost(u) = c_j u^2$ and linear function $Cost(u) = c_j u$.

The probability of responding correctly is a function of the amount of control $u$ that a subject allocates to a given task. We assume that this probability increases monotonically with the amount of control intensity allocated, following the sigmoid function

$$P(O_{correct}|u,S) = \frac{1}{1+e^{-15u-a_{j,S}}} \quad (7)$$

where $a_{j,S}$ determines the degree of task automaticity: The higher $a_{j,S}$, the easier the task, that is, the less cognitive control is needed to reach the correct outcome. Note that $a_{j,S}$ depends on the task environment $S$, as well as the task proficiency of subject $j$.

Finally, an agent's subjective value can be described as a function of the reward (e.g. monetary compensation) provided for a correct response. Here we assume that the value of the correct outcome simply corresponds to

$$V(O_{correct}) = v_j R(O_{correct}) + b_j \quad (8)$$

where $R(O_{correct})$ is a monetary reward that is provided in the event of a correct response, $v_j$ is the reward sensitivity of the agent and $b_j$ is the baseline value that the agent assigns to correct responses (accuracy bias).

## Estimating Control Costs Under Correct Assumptions

We validated the estimation procedure above for three different functional forms of control costs (exponential, quadratic and linear). For each functional form, we parameterized 100 agents with different control cost parameters $c_j$, ranging from 1 to 4 in 100 equally spaced steps. We fixed parameter values for task automaticity ($a_j = -7.5$), reward sensitivity ($v_j = 1$) and accuracy bias ($b_j = 0$) across all agents.

We applied the estimation procedure above to sample multiple estimates of $Cost(u)$, where each sample was obtained under a different reward condition. That is, we manipulated the amount of reward provided for answering correctly $R(O_{correct})$ from \$0 to \$1 in steps of \$0.01 and assessed control cost estimates for each reward condition based on an agent's task performance. Critically, control cost estimates were obtained assuming perfect knowledge of an agent's task automaticity, reward sensitivity, as well as accuracy bias.

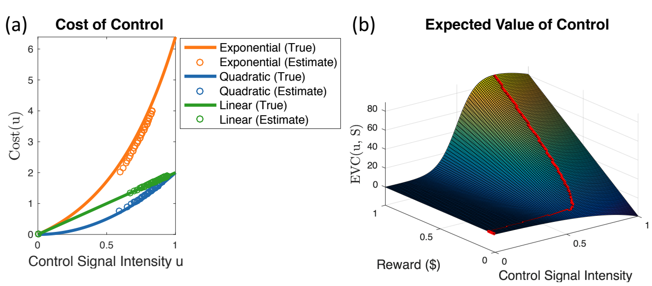


Figure 2: Estimating the cost of cognitive control under correct assumptions. (a) The cost of control as a function of control signal intensity for different functional forms with $c_j = 2$ (solid lines). Circles represent estimates of control costs for a given control signal intensity that were obtained from performance measures taken in different reward conditions. (b) The EVC for an agent with exponential cost function is plotted for varying control signal intensities and reward conditions. The agent chooses to allocate control with an intensity that yields the highest EVC (red line).

We quantified how well the true control cost functions were estimated by first fitting an assumed cost function to sampled control cost estimates. We then linearly regressed agents' true control costs parameter $c_j$ against the estimated parameter $\hat{c}_j$. A perfect control cost estimation, i.e. an identical mapping between true and estimated control costs, should yield a reliable regression coefficient with value 1.

Figure 2a shows estimation results for different functional forms of control costs. Estimated control costs appear to match true control costs for exponential cost functions, $b = 0.99971, t(99) = 3140.1655, p < 10^{-246}$, quadratic cost functions, $b = 0.95311$, $t(99) = 53.8651$, $p < 10^{-73}$, as well as linear cost functions, $b = 0.99518$, $t(99) = 787.3261$, $p < 10^{-187}$. Note that the maximum EVC for non-zero rewards is mostly located at high control signal intensities, limiting the domain for sampled control costs (Figure 2b).

## Estimating the Cost of Cognitive Control Under Incorrect Assumptions

Having provided a proof of concept for the estimation procedure, we will now demonstrate how incorrect assumptions about other motivational variables can lead to potentially drastic misestimations of subjects' control costs. We will first perform a sensitivity analysis that exposes how individual differences in control costs between two simulated agents can be misestimated as a function of systematic individual differences in other variables. We then extend this analysis to investigate how unsystematic variability among other variables can impair the recovery of control costs in a population of agents. Finally, we will show how individual differences in those variables can give rise to spurious correlations between agents' estimated control costs across experiments.

### Sensitivity Analysis

The estimation of individual differences based on task performance relies on assumptions about all variables that contribute to task performance. Here we assess how well we can estimate the true difference in the cost of control for two agents. We are specifically interested in determining boundary cases for which estimates falsely suggest that the true relationship between the control costs for two agents flips.

In this simulation we consider two agents with an exponential cost function, one with a relatively low cost of control ($c_1 = 2$) and one with a relatively high cost of control ($c_2 = 3$). In our estimation of control costs we assume that agents share the same task automaticity ($a'_1 = a'_2 = -7.5$), reward sensitivity ($v'_1 = v'_2 = 1$) as well as the same accuracy bias ($b'_1 = b'_2 = 0$) but we will vary the true values for each parameter away from the assumed value.

Using the estimation procedure described above we attempted to recover the differences in control costs between the two agents as a function of how much the true parameter values for one of the motivational variables deviates from the assumed parameter value. We quantified the relationship between estimated control costs as the difference between estimated control cost parameters for both agents $\hat{c}_2 - \hat{c}_1$. If the distance equals 1 then the relationship is perfectly recovered since $c_2 - c_1 = 1$. If the distance is above 1, or between 0 and 1 then we consider the relationship to be qualitatively in line with the true relationship but quantitatively overestimated or underestimated respectively. A distance below 0 implies a false inference about ordinal differences in costs, i.e. the cost parameter of agent 1 is incorrectly estimated to be larger than the cost parameter of agent 2.

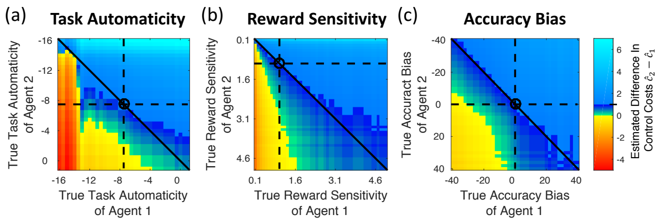


Figure 3: Sensitivity analysis results. Plots indicate the estimated difference in agents control costs for different true values in agents' (a) task automaticity (b) reward sensitivity and (c) accuracy bias. The color indicates the value of the estimated difference in control costs parameters $\hat{c}_2 - \hat{c}_1$. If the distance is 1 then the difference is perfectly recovered. Values below 0 (e.g., yellow and red) indicate reversals of the ordinal relationship of agents' control costs. Values between 0 and 1 (green to dark blue), and greater than 1 (light blue) retain the ordinal relationship but quantitatively under- or overestimate the true difference. Dashed black lines show the assumed parameter value of each agent that is held constant across all simulations. The black diagonal line indicates a fixed ratio between the true parameters of the agents that matches the ratio of assumed parameters.

Figure 3 shows the results of the sensitivity analysis for variations in all agent parameters. As expected, the deviation between the true and assumed values for other motivational parameters can significantly alter the estimated difference in control costs (up to a point at which the ordinal relationship between agents is flipped, see yellow and red areas in Figure 3). However, estimated costs can accurately recover the ordinal relationship of agents' true control costs ($c_1 < c_2$) even if false assumptions are made about other parameters. That is, the cost estimation procedure still recovers a smaller control cost for agent 1 compared to agent 2 ($\hat{c}_1 < \hat{c}_2$) if we underestimate the task automaticity of agent 1 relative to agent 2. Finally, our analysis yields different sensitivity patterns for different agent parameters. For instance, cost estimates appear to be relatively sensitive to changes in the ratio between the agents' task automaticity. However, as the true accuracy bias or reward sensitivity of agent 1 increases, control cost estimates remain robust to deviations in corresponding parameter values for agent 2. Note that these observations are limited to the tested range of true parameter values and the specific set of assumed parameter values.

## Limitations on Control Cost Estimation as a Function of Variability in Other Variables

Several studies have attempted to assess the relationship between proxies for control cost and other, related criteria (e.g. self-control measures) across individuals (Kool et al., 2013; Westbrook et al., 2013; Gold et al., 2015). Often, these approaches do not factor in other motivational variables (e.g. task automaticity, reward sensitivity or accuracy bias), thus making the implicit assumption that participants don't differ with respect to these variables. Here we extend our previous analysis to investigate how unsystematic variability among agents in terms other motivational variables can impose constraints on the ability to recover control costs. Critically, any such constraint would limit an experimenter's ability to reveal potential relationships between the true control cost of a subject and other, related criteria (e.g. self-control measures).

We performed separate investigations for each motivational variable. Every investigation entailed separate experiments, each involving 100 simulated agents. The control cost parameter $c_j$ was varied across agents from 1 to 4 in steps of 0.02 within an experiment. The parameter for the motivational variable of interest was drawn from a normal distribution with means $\mu_a = 7.5$ (task automaticity), $\mu_v = 1$ (reward sensitivity) and $\mu_b = 0$ (accuracy bias). The standard deviation for the relevant variable was varied across experiments while the other two motivational variables were fixed to the mean. The standard deviation was varied for all motivational variables $\sigma_a^2, \sigma_b^2, \sigma_v^2$ from 0 to 10 in steps of 1. Each standard deviation condition involved 10 experiments.

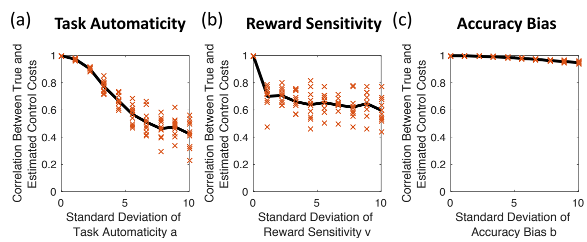


Figure 4: Recovery of control costs as a function of variability in other motivational variables. The correlation between true and estimated control costs within an agent pool is shown as a function of variation in either (a) task automaticity, (b) reward sensitivity or (c) accuracy bias across agents. Solid lines plot the mean correlation. Each cross corresponds to a simulation experiment with 100 agents.

We obtained control cost estimates for each experiment by varying the reward of the correct outcome $R(O_{correct})$ from \$0 to \$1 in steps of \$0.01 and applying the estimation procedure described above under the (false) assumption that parameters for motivational variables are the same across agents ($a'_j = -7.5, v'_j = 1, b'_j = 0$). We then assessed our ability to recover control cost estimates, quantified as the correlation between true and estimated control cost parameters across agents. To investigate how this correlation depends on the variability for any given motivational variable in the agent pool, we linearly regressed the correlation between true and estimated costs against the standard deviation of that variable.

We find that the correlation between true and estimated control costs decreases with an increased population standard deviation of task automaticity, $b = -0.064$, $t(99) =$

$-25.9268$, $p < 10^{-44}$, reward sensitivity, $b = -0.0243$, $t(99) = -7.12$, $p < 10^{-9}$, and accuracy bias, $b = -0.0053$, $t(99) = -35.1012$, $p < 10^{-56}$. That is, the presence of (unaccounted) variation in any of those variables among subjects can limit the experimenter's ability to estimate true individual differences in control costs. Interestingly, the spread of correlations across experiments increases with the variance of task automaticity and reward sensitivity, suggesting that the reliability of cost estimates should decrease as cross-subject variance in those variables increases.

### Spurious Correlations of Control Cost Estimates Between Different Experiments

Apart from studying differences in subjects' control costs within an experiment, a researcher may be interested in relating such estimations across experiments, for instance, by correlating different proxies for control costs, e.g. measures of demand avoidance in the demand selection task (DST, Kool et al., 2010) with measures of cognitive effort discounting (COGED Westbrook et al., 2013). However, the cognitive control mechanisms involved in the two tasks may not be the same. Thus, it is possible that the costs of control across the two paradigms are not related. Here we will explore whether correlations between estimated control costs across paradigms can arise despite the absence of a true correlation.

Similar to the previous section we investigate the effect of each motivational variable (task automaticity, reward sensitivity, accuracy bias) separately. Each scenario involves 100 agents that are tested in two different paradigms, each of which may require different types of control signals. To expose spurious correlations we will assume that an agent's true cost parameter in paradigm 1 is unrelated to its true cost parameter in paradigm 2. Thus, any estimated non-zero correlation of agents' control costs between paradigms would be spurious. However, we randomly sampled the parameter for the motivational variable of interest such that there was a correlation between the true parameter value in paradigm 1 and the corresponding parameter value in paradigm 2 ($r_a$, $r_v$, $r_b$). E.g., a high correlation of task automaticity across paradigms would imply that an agent with higher task automaticity in paradigm 1 would also have a higher task automaticity in paradigm 2. We varied this correlation from 0 to 1 in steps of 0.1 across experiments. Other motivational variables were held constant ($a_j = -7.5$, $v_j = 1$, $b_j = 0$).

To estimate the control cost parameters we varied $R(O_{correct})$ across experiment conditions and applied the estimation procedure described above. We assessed the correlation $r_{\hat{c}}$ between the estimated control cost parameters of both paradigms for each experiment. This correlation was taken as the dependent variable and regressed against the manipulated correlation of the motivational variable $r_a$, $r_v$ or $r_b$.

Our results yield spurious correlations between control costs estimated across different paradigms. These spurious correlations increased as the relationship between agents' task automaticity, $b = 0.4670$, $t(99) = 15.9403$, $p < 10^{-29}$ and reward sensitivity, $b = 0.2881, t(99) = 9.4694, p < 10^{-14}$ increased across paradigms. However, no systematic spurious correlations occurred for cross-experiment correlations in accuracy bias, $b = -0.0195$, $t(99) = -0.6204$, $p = 0.5366$.

## General Discussion and Conclusion

The cost of cognitive control and its estimation from behavior has become an attracting field of study for many researchers. While some studies were able to identify relationships between behavioral proxies of control costs and other criteria, they were often derived without theory and under neglect of other, confounding variables. In this work we derived a cost estimation method from a computational model of control allocation (Shenhav et al., 2013; Musslick et al., 2015). We demonstrated that it can recover the functional form and parameterization of the cost of control from task performance.

Yet, our results reveal how quantitative and qualitative misestimations can arise if the experimenter does not take into account individual differences related to other variables (e.g. task automaticity, reward sensitivity and, to a lesser extent, accuracy bias) that confound the measures from which control costs are estimated. The sensitivity analyses described here provide a way of assessing the strength of these distortions. However, the analysis revealed that it is possible to correctly estimate ordinal relationships of individual costs, especially if one makes correct assumptions about the ordinal relationship of other motivational parameters between subjects. Moreover, true ordinal relationship between the control costs of two subjects (e.g. subject A having a higher cost than subject B) is more likely to be identified if these two subjects share a similar relationship with respect to other variables (e.g. subject A has a lower reward sensitivity, lower accuracy bias and lower task automaticity than subject B). This suggests that the experimenter may not need perfect knowledge about the exact value for such parameters (e.g. task automaticity) for every subject if, instead, the experimenter knows about the ordinal relationship of those parameters between subjects. Finally, we demonstrated that any between-subject variability in other motivational variables can (a) generally limit the ability to recover true costs and (b) lead to spurious correlations of proxies for control costs across experiments. Accounting for these individual differences is therefore critical to obtain valid cost estimates from individuals – a practice that has been largely neglected in previous work.

These results have significant implications for attempts to estimate individual differences in the cost of control from behavioral measures. That is, the experimenter must ask: If a person shows more effort avoidance or lower task performance , is it because they are less practiced, less motivated or because they have a higher cost of control? To answer these questions we recommend that researchers take additional assessments of these variables into account when estimating individual differences. This may involve simple metrics such as surveys about socioeconomic status or neural correlates of sensitivty to rewards that are not tied to performance. Experimental manipulations may be able to reduce variance in confounding measures, such as extensive training on a task to achieve a similar level of task automaticity across subjects.

Several other factors should be taken into account when considering results from the presented estimation procedure. The estimation procedure was performed on a noise-free, continuous measure of performance (task accuracy) under perfect measurement conditions. We also did not incorporate other biases into our analysis that may confound estimates of control costs from task performance, such as intrinsic rewards associated with control allocation (Inzlicht, Shenhav, & Olivola, 2018). Our results may therefore depict a rather liberal view of the limitations associated with estimating control costs under false assumptions. Future extensions of this work will consider alternative expected value formulations, such as reward rate, as well control cost estimates from other behavioral measures like task choice.

A promising step towards a reliable estimation of control costs is an improved understanding about why these costs exist in the first place. Recent work suggests that these costs reflect forgone opportunities for controlled processing that result from the inability to carry out multiple controlled processes at the same time (Kurzban et al., 2013). The latter can be attributed to a fundamental tradeoff in neural systems between the learning benefit that is gained from shared representations and the bottlenecks shared representations incur for multitasking (Feng, Schwemmer, Gershman, & Cohen, 2014; Musslick et al., 2016, 2017). From this view, control costs may serve to reduce interference, by limiting the number of controlled processes engaged at the same time. This suggests an unexplored possibility for estimating control costs based on the degree of shared representation between tasks – a measurement that is independent of performance.

In conclusion, we argue that model-based estimation procedures can significantly improve our understanding about the validity of estimates of the cost of control, by revealing the conditions for reliable estimation. We hope that the insights gained from our analysis will help to yield more reliable and valid individual difference studies on the cost of cognitive control.

## References


Barch, D. M., Treadway, M. T., & Schoen, N. (2014). Effort, anhedonia, and function in schizophrenia: reduced effort allocation predicts amotivation and functional impairment. *Journal of abnormal psychology*, *123*(2), 387.

Botvinick, M., & Braver, T. (2015). Motivation and cognitive control: from behavior to neural mechanism. *Annual Review of Psychology*, *66*.

Feng, S. F., Schwemmer, M., Gershman, S. J., & Cohen, J. D. (2014). Multitasking vs. multiplexing: toward a normative account of limitations in the simultaneous execution of control-demanding behaviors. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(1), 129-146.

Gold, J. M., Kool, W., Botvinick, M. M., Hubzin, L., August, S., & Waltz, J. A. (2015). Cognitive effort avoidance and detection in people with schizophrenia. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(1), 145–154.

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

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.

Kool, W., McGuire, J. T., Wang, G. J., & Botvinick, M. M. (2013). Neural and behavioral evidence for an intrinsic cost of self-control. *PloS one*, *8*(8), e72626.

Krebs, R. M., Boehler, C. N., & Woldorff, M. G. (2010). The influence of reward associations on conflict processing in the stroop task. *Cognition*, *117*(3), 341–347.

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), 661–679.

Musslick, S., Dey, B., Özcimder, K., Patwary, M. M. A., 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 conference of the Cognitive Science Society* (pp. 1547–1552). Philadelphia, PA.

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 conference 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 *The 2nd multidisciplinary conference on reinforcement learning and decision making.*

Padmala, S., & Pessoa, L. (2011). Reward reduces conflict by enhancing attentional control and biasing visual cortical processing. *Journal of cognitive neuroscience*, *23*(11), 3419–3432.

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.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual review of neuroscience*, *40*, 99–124.

Umemoto, A., & Holroyd, C. B. (2015). Task-specific effects of reward on task switching. *Psychological research*, *79*(4), 698–707.

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

Westbrook, A., Kester, D., & Braver, T. S. (2013). What is the subjective cost of cognitive effort? load, trait, and aging effects revealed by economic preference. *PLoS One*, *8*(7), e68210.