Decomposing Individual Differences in Cognitive Control: A Model-Based Approach

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

[1]Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA.
[2]Department of Cognitive, Linguistic, and Psychological Sciences,
Brown Institute for Brain Science, Brown University, Providence, RI 02912, USA.

[*]Corresponding Author: *musslick@princeton.edu*

Abstract

Researchers have long been interested in using laboratory measures of cognitive control to predict a person's cognitive control/self control success outside the lab. We used a computational approach to identify which lab-based performance measures provide the most valid individual difference measures of one's ability and/or motivation to exert cognitive control. We simulated performance across an array of cognitive control tasks, and estimated the degree to which different performance metrics (e.g., congruency effects, conflict adaptation, and demand avoidance) could theoretically provide valid estimates of processes underlying control allocation. By performing dimension reduction on these performance metrics, we further revealed latent dimensions that can index separate mechanisms of control-demanding behavior. Our results suggest that individual differences in measures of cognitive control can originate from multiple factors, several of which are unrelated to *capacity* for cognitive control. We conclude by discussing implications of these analyses for assessing individual differences in cognitive control phenomena.

Keywords: individual differences; cognitive control; motivation; self-control

Introduction

Cognitive control refers to our ability to adapt mental processes to current task goals. Researchers have developed a variety of measures to index a given person's capacity to exert cognitive control, such as conflict-related interference, conflict adaption, and performance costs associated with task switching. It has often been assumed that individual differences in capacity and/or motivation for control should predict one's self-control success in the real world, and that performance on one cognitive control task should therefore correlate with indices of self-control. Unfortunately, however, such correlations have been inconsistent across the literature. For instance, whereas some individual studies find correlations between Stroop conflict-related interference (congruency costs) and real-world self-control outcomes (e.g., addiction treatment compliance, healthy diets; Streeter et al., 2008; Allan, Johnston, & Campbell, 2010), a large study (N=2,641) recently found no correlation between congruency costs and a well-validated index of real-world self-control (Saunders, Milyavskaya, Etz, Randles, & Inzlicht, 2017). The inconsistency in these findings has been taken to suggest that control mechanisms are highly context-specific and/or that self-control may not actually require cognitive control (Berkman, Hutcherson, Livingston, Kahn, & Inzlicht, 2017). Here we explore an alternative interpretation, that commonly used measures of control allocation may be ill-suited to indexing the control required of those tasks.

Converging evidence suggests that performance on cognitively demanding tasks reflects a combination of bottom-up stimulus processing and one's capacity and motivation to exert top-down control over such processing (Cohen, Dunbar, & McClelland, 1990; Shenhav, Botvinick, & Cohen, 2013; Shenhav et al., 2017). These insights have been integrated into a recent computational model of control allocation, which simulates an agent's performance on a cognitive task based on the parameters of that task (e.g., stimulus salience) and the incentives on offer (e.g., reward for correct response; Musslick, Shenhav, Botvinick, & Cohen, 2015). Control allocation is determined by comparing the expected reward (based on the incentives and the degree to which control increases the likelihood of a correct response) with an intrinsic cost of control, to determine the overall Expected Value of Control (EVC). The parameters of these simulated agents can be adjusted to vary how they process stimuli and incentives (e.g., their sensitivity to rewards), resulting in attendant changes to task performance. Theoretical analyses suggest that between-subject variability in some motivational parameters, such as reward sensitivity, can generally limit the ability to recover other motivational parameters, such as the cost of cognitive control, from task performance (Musslick, Cohen, & Shenhav, 2018; Caplin, Csaba, Leahy, & Nov, 2018). An important question that remains unaddressed, however, is whether individual differences in cognitive control phenomena provide a reliable index for one's capacity to exert cognitive control.

Here, we use the EVC model to simulate various phenomena that have been used to index one's capacity to exert cognitive control, including within-trial interference and cross-trial adaptation to response conflict; task-switching costs; and cognitive effort discounting. We then demonstrate that individual differences in these phenomena are influenced by parameters of the task and the agent, including variables related to bottom-up stimulus processing, the *ability* to exert control, and the *motivation* for doing so. Finally, we identify latent dimensions that explain individual differences across these simulated phenomena, and discuss implications of this work for the assessment of individual differences in cognitive control within and outside of the lab.

Expected Value of Control Model

The EVC theory is based on the premise that control allocation involves specifying the identity of candidate control signals, as well as the intensity of each (Shenhav et al., 2013). Increases in control signal intensity lead to improvements in performance on the corresponding task. However, it is also assumed that exercising cognitive control is costly and this cost increases monotonically with the intensity of the control signal. According to the EVC theory, the control system chooses to implement the configuration of control signals that yields the highest expected value of control, that is, the expected utility of implementing a configuration of control signals with specified intensities minus their associated costs. Critically, the expected value for each candidate control signal configuration is contingent on an internal model of the task environment that is updated based on experience.

The present implementation of the EVC model describes performance in the Stroop task (e.g., responding to the ink color of a color word, Stroop, 1935), in terms of an interaction between the control system and the task environment. The control signal is chosen optimally based on an internal model of the next trial which produces an estimate of the next trial (inferred state $\hat{\mathbf{S}}$). This signal is then used to interact with the environment (actual state $\mathbf{S}$), for example to commit one of the two possible responses[1] in the task. After each trial, the agent updates the internal model based on an observation of that trial following the response.

In order to generate reaction times (RTs) and responses on each trial, we use the drift diffusion model (DDM Ratcliff, 1978). Within the DDM framework, a response on the task can be conceptualized as a result of the noisy accumulation of evidence toward one of the two possible responses (e.g. one response indicating the color green and the other response indicating the color red; Musslick et al., 2015). Here, we assume that the rate of evidence accumulation toward one of the two responses is governed by a controlled and an automatic component

$$drift = \varepsilon \cdot drift_{\text{control}} + drift_{\text{automatic}} \tag{1}$$

where ε is a capacity parameter that scales the amount of control allocated. The automatic component reflects automatic processing of the color feature and word feature of the stimulus that is unaffected by control,

$$drift_{\text{automatic}} = a_{\text{color}} + a_{\text{word}}. \tag{2}$$

The absolute magnitude of the color-response association a_{color}, as well as the magnitude of the word-response association a_{word} depends on the strength of the association of each stimulus feature with a given response, and its sign depends on the response (e.g. $a_{\text{color}} < 0$ if the response is associated with the left button, $a_{\text{color}} > 0$ if response is associated with the right button). Thus, for congruent trials a_{color}, and a_{word} have the same sign, whereas the opposite sign for incongruent trials. The controlled component of the drift rate is the sum of the two stimulus values, as well as the intensity of the corresponding control signal, one for processing the color dimension of the stimulus u_{color} and one for processing the word dimension of the stimulus u_{word}:

$$drift_{\text{control}} = u_{\text{color}} \cdot a_{\text{color}} + u_{\text{word}} \cdot a_{\text{word}} \tag{3}$$

Thus, each control signal biases processing towards one of the two stimulus dimensions, both of which characterize the actual state on a given trial, $\mathbf{S} = \{a_{\text{color}}, a_{\text{word}}\}$. As a result, higher control signal intensity for processing the color dimension improves performance — speeds responses and lowers error rates — in a trial of the Stroop task. Mean RTs and response probabilities for a given parameterization of drift rate on trial t are derived from an analytical solution to the DDM (Navarro & Fuss, 2009).

In order to specify the optimal set of control signals $\mathbf{U} = \{u_{\text{color}}, u_{\text{word}}\}$ on a given trial t, the model estimates the expected value for each configuration of control signal intensities based on its internal model of the next trial $\hat{\mathbf{S}} = \{\hat{a}_{\text{color}}, \hat{a}_{\text{word}}\}$. This is done by weighting the expected reward for an outcome against the cost associated with the chosen control signal configuration:

$$EVC(\mathbf{U}, \hat{\mathbf{S}}) = P(\text{correct}|\mathbf{U}, \hat{\mathbf{S}})V(R) - Cost(\mathbf{U}) \tag{4}$$

where $P(\text{correct}|\mathbf{U}, \hat{\mathbf{S}})$ corresponds to the probability of reaching the decision threshold for the correct response and $V(R)$ corresponds to the subjective value of responding correctly. Here, the subjective value $V(R) = vR$ corresponds to the amount of reward offered for a correct response R weighted by the model's sensitivity to the reward v. The cost $Cost(\mathbf{U}) = Cost_{\text{impl}}(\mathbf{U}) + Cost_{\text{reconf}}(\mathbf{U})$ is composed of an implementation cost that increases with the amount of control being allocated (Shenhav et al., 2013; Manohar et al., 2015; Lieder, Shenhav, Musslick, & Griffiths, 2018),

$$Cost_{\text{impl}}(\mathbf{U}) = e^{c_{\text{I}} \cdot u_{\text{color}}} + e^{c_{\text{I}} \cdot u_{\text{word}}} \tag{5}$$

as well as a reconfiguration cost that scales with the degree to which control signals need to be changed relative to their previous state (Meiran, 1996; Rogers & Monsell, 1995)

$$Cost_{\text{reconf}}(\mathbf{U}) = e^{c_{\text{R}} \sqrt{(u_{\text{color},t} - u_{\text{color},t-1})^2 + (u_{\text{word},t} - u_{\text{word},t-1})^2}} \tag{6}$$

where the implementation cost is scaled by parameter c_{I} and the reconfiguration cost is scaled by parameter c_{R}. The model selects the control signal configuration with the maximum EVC within the inferred next trial $\hat{\mathbf{S}}$, out of all the configurations under consideration:

$$\mathbf{U}^* = \underset{\mathbf{U}}{\text{argmax}} EVC(\mathbf{U}, \hat{\mathbf{S}}) \tag{7}$$

Performance in the actual state $\mathbf{S}$ is determined by the influence of the chosen control signals on the true parameters

[1] A restriction to two response alternatives limits the scope of the model to paradigms with two-alternative forced choice but makes it amenable to tractable computation of mean reaction times and error rates.

a_{color} and a_{word}. After observing the actual state, the agent updates its inferred state $\hat{\mathbf{S}} = \{\hat{a}_{\text{color}}, \hat{a}_{\text{word}}\}$:

$$\hat{a}_{\text{color, new}} = \hat{a}_{\text{color, old}} + \alpha(\hat{a}_{\text{color, old}} - a_{\text{color}}) \quad (8)$$

$$\hat{a}_{\text{word, new}} = \hat{a}_{\text{word, old}} + \alpha(\hat{a}_{\text{word, old}} - a_{\text{word}}) \quad (9)$$

where α is the learning rate. Finally, the agent re-evaluates the optimal control policy for the next trial based on its revised model of the task environment.

Task Environments and Parameterization

We simulate behavior of the EVC agent across three different experimental paradigms that have been repeatedly used to index individual differences in cognitive control. Here, we describe each paradigm, the associated behavioral phenomena, as well as the corresponding parameterization[2] of the EVC model.

Stroop Task

In the Stroop paradigm, the agent is presented with a two-dimensional stimulus, one dimension representing an ink color and another dimension representing a color word (Stroop, 1935). On each trial, the EVC model is required to indicate the response associated with the ink color. In congruent trials, the word feature of the stimulus is associated with the same response as the ink color whereas in incongruent trials, the color and word features are associated with different responses. The experiment sequence encompassed 101 trials, and was fully balanced (excluding the first trial) with respect to congruent and incongruent stimuli, as well as with respect to all four transitions between the two trial types (congruent-congruent, congruent-incongruent, incongruent-congruent, incongruent-incongruent). As described below, we sampled a_{color} uniformly from $U(0.3, 0.4)$. To simulate congruent trials, we set $a_{\text{word}} = 0.4$ such that both stimuli dimensions promote the same response. On incongruent trials, we set $a_{\text{word}} = -0.4$ such that the word dimension is associated with a different response than the color dimension. Note that the absolute magnitude of a_{word} is higher than a_{color}, reflecting the assumption that word reading is a more automatic process than color naming (Cohen et al., 1990). We varied the range of control signal intensities from 0 to 10 in steps of 0.2 for the two control signals u_{color}, u_{word} and set the reward received for a correct response to $R = 100$. DDM parameters were set as follows: starting point = 0.0, noise coefficient = 0.7, non-decision time = 0.2s and threshold = 0.4.

We used this paradigm to simulate three different behavioral phenomena. One of the most reliable observations is that participants take more time and commit more errors when responding to incongruent stimuli as opposed to congruent stimuli (Stroop, 1935). Here, we assessed effects of stimulus congruency as the difference in RTs and error rates between incongruent and congruent trials. Another common observation is that participants exhibit a smaller performance cost for incongruent stimuli when the current stimulus was preceded by an incongruent stimulus as opposed to a congruent stimulus (Gratton, Coles, & Donchin, 1992; Egner, 2007). We assessed the congruency sequence effect as an interactive effect between the congruency of the current trial and the congruency of the previous trial on performance. Finally, participants tend to exert smaller congruency effects when the proportion of congruent stimuli is decreased (proportion congruency effect, Logan & Zbrodoff, 1979). We assessed this phenomenon by comparing the congruency effect in two different experiment sequences, one that contained 20% congruent trials, and one that contained 80% congruent trials.

Task Switching

The performance costs associated with switching from one task to another are often used to index cognitive flexibility (Koch, Poljac, Müller, & Kiesel, 2018; Rogers & Monsell, 1995). Here, we examined this effect in a cued task switching paradigm in which the model had to switch between categorizing the color of a stimulus (color naming) and categorizing its shape (shape naming). Similar to the Stroop task, stimuli were either congruent, $a_{\text{color}} = a_{\text{shape}}$, or incongruent, $a_{\text{color}} = -a_{\text{shape}}$. The trial sequence encompassed 100 trials that were randomly sampled with respect to stimulus congruency (congruent, incongruent), the currently relevant task (color naming, shape naming) and the task transition with respect to the previous trial (task switch, task repetition). On each trial, the model allocated control between the two control signals u_{color}, u_{shape}, using the same range of control intensities as described in the Stroop task. The model was cued with a baseline reward of $R = 100$, providing information about which feature is relevant for the task it has to perform on the current trial. DDM parameters were set as follows: starting point = 0.0, noise coefficient = 0.3, non-decision time = 0.2s and threshold = 0.15.

We assessed switch costs in terms of the difference in RTs and error rates between task switch trials and task repetition trials. Rogers and Monsell (1995) also demonstrated that congruency costs are higher on task switch trials compared to task repetition trials. To capture this effect, we also assessed the interaction between stimulus congruency and task transition.

Cognitive Effort Discounting

When given a choice between performing a task with low cognitive effort and a task with high cognitive effort, participants tend to select the former, even if it means to forgo a reward (Westbrook & Braver, 2015). Here, we simulated demand avoidance in the cognitive effort discounting (COGED) experiment described by Westbrook and Braver (2015). In this paradigm, subjects can choose on each trial whether they want to perform a baseline low-demand task for a low reward or a higher-demand alternative task for a higher reward. The amount of reward offered for the baseline task is adjusted to

[2]Note that fixed parameters for each paradigm were chosen such that the model performed with at least 55% accuracy for all combinations of individual difference parameters.

identify the point of indifference, that is, the reward at which subjects are indifferent between performing the low-demand baseline task and performing the high-demand task. To simulate this paradigm, we modeled both tasks as different types of trials that the model can choose between. Each trial encompassed a stimulus with a color dimension that mapped to one of two responses with $a_{\text{color}} > 0$. However, unlike in the Stroop task there was no word dimension, $a_{\text{word}} = 0$. The difficulty of the high-demand task was manipulated across experiment blocks, by varying the color-response association a_{color} from 1.0 to 0.2 in steps of 0.2, and the difficulty of the baseline task was fixed to $a_{\text{color}} = 1$ (higher color-response associations may reflect higher saturation values for a color patch). For each set of simulations, we fixed the reward for the high-demand task to $R = 200$ while steadily increasing the amount of reward offered for the low-demand task in steps of 1, beginning from an initial reward value of $R = 1$. On each trial, the EVC agent determined the highest EVC separately for each task and chose the task with the highest predicted EVC. We then assessed the amount of reward offered for the low-demand task for which the model would be indifferent between performing the low-demand task and the (more rewarding) high-demand task, and normalized this value by the amount of reward offered for the high-demand task. Following the notation by Westbrook and Braver (2015), we refer to this normalized value as the subjective value of completing the high-demand task. For instance, if the model would switch to performing the low-demand task at an offered reward of 120 then the (discounted) subjective value of the high-demand task would be 120/200. The range of control signal intensities was varied from 0 to 10 in steps of 0.2 and DDM parameters were set as follows: starting point = 0.0, noise coefficient = 1.5, non-decision time = 0.2s and threshold = 1. We assessed subjective value the high-demand task as a function of its difficulty, $1 - a_{\text{color}}$.

Simulation Procedure

We simulated behavior of 100 EVC agents in the three paradigms described above. For each agent, we uniformly sampled its control capacity $\varepsilon \sim U(0.5, 1.5)$, implementation cost $c_I \sim U(0.5, 1.5)$, reconfiguration cost $c_R \sim U(0, 3)$, reward sensitivity $v \sim U(0.5, 1)$, the stimulus-response association of the relevant task ($a_{\text{color}} \sim U(0.3, 0.4)$ in all paradigms[3], as well as a_{shape} in the task switching paradigm) and learning rate $\alpha \sim U(0, 0.5)$. Ranges for these parameters were chosen to warrant an accuracy above 55% across all simulated paradigms. Note that agents with a higher control capacity would effectively implement a higher amount of control. Therefore, control capacity can be taken as a proxy for the amount of control an agent exerts on average. The stimulus-response association determines the degree of task automaticity: The higher the stimulus-response association of a task-relevant feature, the easier the task, that is, the less cognitive control is needed to reach the correct outcome. Here, we assume that the stimulus-response association of a task feature reflects the task proficiency of an agent.

We first assessed average behavior across all agents with respect to seven dependent variables. In the Stroop task, we measured error rate effects of stimulus congruency, the congruency sequence effect, the proportion congruency effect, as well as overall error rate on the task. In the task switching paradigm, we assessed switch costs in error rates, as well as the congruency costs in error rates as a function of task transition. We also measured the subjective value of levels of task difficulty as determined by the COGED paradigm.

We restricted our analysis of individual differences to overall error rate in the Stroop task, congruency effects, congruency sequence effects, proportion congruency effects, switch costs, as well as the subjective value assigned to a task parameterized with a_{color} (effort discounting). We then took two different approaches to analyze individual differences in these measures. First, we used a multiple linear regression to assess the degree to which each of the six EVC parameters can explain each behavioral phenomenon. However, we did not include learning rate as a regressor in the task switching and COGED paradigms as the agent is provided full information about each trial. Second, we used principal component analysis (PCA) to explore whether individual differences can be explained by more complex latent factors. That is, we identified principal components that account for variance between agents (observations) across all dependent variables (dimensions), including overall error rate, congruency effect, congruency sequence effect, proportion congruency effect, switch cost and effort discounting. We then assigned a score to each agent that identifies its position on the axes spanned by either the first or the second principal component. These two components explain most of the variance in the space of behavioral phenomena, and can be best interpreted in terms of the behavioral effects that vary most along a given component. In addition, we sought to interpret each component in terms of individual difference parameters of the EVC model. That is, we identified the individual difference parameters that best explain each principal component, by regressing the component scores of all agents against their EVC parameters. Finally, we assessed which of the behavioral phenomena were most indicative of the amount of exerted control, by computing the Pearson correlation between each dependent variable (e.g. congruency effect) and the average intensity of control u that an agent exerts, across all agents.

Results

Behavioral Phenomena. The EVC model captured all of the cognitive control phenomena of interest[4] (Figure 1): 1) Responses were slower and more error-prone on incongruent versus congruent trials of a Stroop-like task (*congruency effect*), $F(1, 99) = 17.80$, $p < 0.001$. 2) When the stimuli on

[3] In the COGED task, we scaled the tested range of a_{color} by this value.

[4] We focused our analyses on error rates due to space constraints. However, we observed similar effects for RTs.

Table 1: Regression of behavioral phenomena against individual differences in EVC parameters. Significant regressors are ordered by standarized regression weight.

Model Parameter	β	t	p
Overall Error Rate, $df = 93$			
Task Automaticity	-0.631	-5.48	< 0.001
Control Capacity	-0.127	-11.77	< 0.001
Implementation Cost	0.126	11.32	< 0.001
Learning Rate	-0.114	-5.05	< 0.001
Reward Sensitivity	-0.076	-3.63	< 0.001
Congruency Effect, $df = 93$			
Task Automaticity	-0.710	-3.50	< 0.001
Learning Rate	-0.219	-5.50	< 0.001
Implementation Cost	0.114	5.82	< 0.001
Control Capacity	-0.089	-4.69	< 0.001
Congr. Sequence Effect, $df = 93$			
Learning Rate	0.145	6.29	< 0.001
Reward Sensitivity	0.055	2.53	< 0.05
Control Capacity	0.053	4.78	< 0.001
Implementation Cost	-0.031	-2.71	< 0.01
Reconfiguration Cost	-0.031	-7.70	< 0.001
Proportion Congr. Effect, $df = 93$			
Task Automaticity	-0.471	-2.19	< 0.05
Learning Rate	-0.199	-4.72	< 0.001
Reward Sensitivity	0.160	4.08	< 0.001
Control Capacity	0.112	5.56	< 0.001
Implementation Cost	-0.103	-4.92	< 0.001
Reconfiguration Cost	-0.044	-6.04	< 0.001
Switch Cost, $df = 94$			
Implementation Cost	-0.069	-4.42	< 0.001
Reward Sensitivity	0.059	2.00	< 0.05
Control Capacity	0.038	2.49	< 0.05
Reconfiguration Cost	0.015	2.77	< 0.01
Effort Discounting, $df = 88$			
Task Automaticity	-0.603	-7.77	< 0.001
Implementation Cost	0.139	17.44	< 0.001
Control Capacity	-0.051	-6.86	< 0.001
Reconfiguration Cost	-0.047	-17.48	< 0.001

the previous trial were incongruent, congruency effects were smaller on the current trial, relative to when the previous trial was congruent (*congruency sequence effect* or *conflict adaptation*), $t(99) = 4.22$, $p < 0.001$. 3) Congruency effects were higher when the trial sequence contained a high proportion of congruent trials versus a high proportion of incongruent trials (*proportion congruency effect*), $t(99) = 17.86$, $p < 0.001$. 4) Responses were less accurate when switching to a new task rather than repeating the same task (*switch costs*, Rogers & Monsell, 1995), $F(1,99) = 337.30$, $p < 0.001$. These switch costs were greater when transitioning to an incongruent trial (Rogers & Monsell, 1995), $F(1,99) = 214.96$, $p < 0.001$. 5) All else being equal, simulated agents assign less value to (and would therefore be less likely to engage with) tasks that are more rather than less difficult (*cognitive effort discounting*, see Figure 1D).

Individual Differences. We tested the degree to which each of the measures above were influenced by individual differences in factors related to bottom-up stimulus processing (task automaticity), cognitive control ability (control capacity), and motivational factors (e.g., reward sensitivity and control costs). Agents with a higher control capacity and lower implementation costs made fewer errors, had lower congruency effects, higher congruency sequence effects, adapted more to the proportion of congruent trials, had higher switch costs and discounted cognitive effort less (Table 1). Agents with higher reconfiguration costs and a lower sensitivity to reward adapted less to congruency of the previous stimulus or to the proportion of congruent trials. Both, a higher reconfiguration cost and a higher reward sensitivity were associated with higher switch costs. A higher reward sensitivity also yielded overall fewer errors while higher reconfiguration costs predicted less effort discounting. Agents with a higher learning rate and task automaticity performed overall better in the Stroop task, showing smaller congruency effects and smaller proportion congruency effects. Unsurprisingly, agents with a higher learning rate show greater sequential adaptations to response congruency whereas agents with a higher task automaticity discounted effort less.

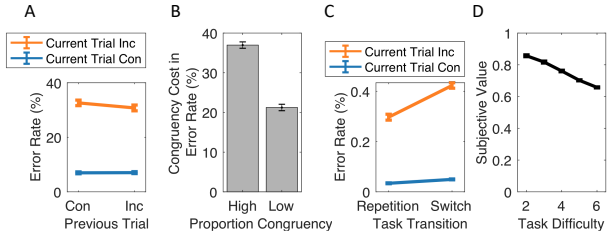

Figure 1: Average effects of simulated agents. (A) Error rates are shown as a function of congruency of the previous and the current trial. (B) Congruency effects (difference in error rates on incongruent and congruent trials) are shown for a sequence with a low (80%) and high (20%) proportion of congruent trials. (C) Error rates are shown as a function of congruency of the previous trial and task transition. (D) Subjective value of a task as a function of its difficulty. Error bars indicate the standard error of the mean across simulated agents.

Principal Components Analysis. After performing a PCA across our behavioral effects of interest, we found that individual differences across these are well captured by two orthogonal dimensions that explained more than 75% of between-agent variance (Figure 2). Regressing these phenomenon-driven components on the model parameters that we varied, we find that a high score on Component 1 is associated with higher task automaticity, lower implementation costs, higher control capacity and higher sensitivity to reward. Agents with a higher value for any of these parameters are expected to perform better on a task (Table 3). Component 2 appears to most reliably capture differences in reconfiguration costs, and to a lesser degree differences in task automaticity, reward sensitivity and implementation costs.

Correlation with control intensity. We found that each behavioral effect significantly correlated with the average amount of control exerted by an agent (Table 2). Interestingly, overall error rate in the Stroop task was most indicative of exerted control intensity, followed by incongruency costs.

General Discussion and Conclusion

People have varying degrees of success at adapting their thoughts and behaviors to meet their current goals. Failing

Table 2: Correlations between dependent behavioral measures and exerted control intensity across simulated agents ($df = 98$).

Dependent Measure	r	p
Overall Error Rate	-0.76	< 0.001
Congruency Effect	-0.67	< 0.001
Proportion Congruency Effect	0.58	< 0.001
Effort Discounting	0.46	< 0.001
Congruency. Sequence Effect	0.31	< 0.01
Switch Cost	0.20	< 0.05

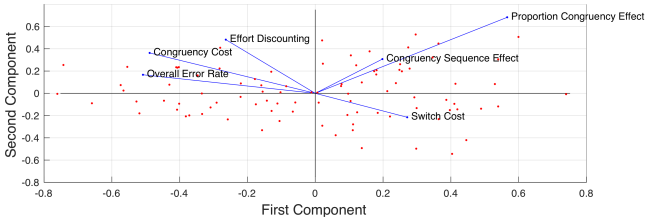

Figure 2: Principal Components Analysis. Each red dot summarizes the behavior of an agent in the space of the first and second principal component. The direction and length of the blue vectors indicates the score of each behavioral effect in terms of the two components. For instance, subjects with low scores on the first component appear to commit more errors but show lower costs of switching tasks.

to exert the appropriate level of control can have very negative consequences for one's health, career, and social status. It is therefore important to understand whether and how such real-world self-control can be predicted from lab-based measures of cognitive control. We used a computational model of control allocation to examine the degree to which different performance metrics from such tasks can theoretically index individual differences in processes related to stimulus processing, strength of control, and motivation for control.

We showed that the EVC model can account for a wide array of effects used to index cognitive control, including response interference, sequential adaptation to stimulus congruency, adaptation to the proportion of congruent stimuli, performance costs associated with task switching, and demand avoidance. Critically, we showed that individual differences in each of these measures can be accounted for by a multitude of factors, including motivational variables (e.g., reward sensitivity) and bottom-up stimulus processing (task automaticity), rather than only by one's *ability* to exert cognitive control (indexed by control capacity). This suggests that individual differences in cognitive control phenomena may not not be a reliable indicator of one's ability to exert control but may instead reflect individual differences in other variables. A PCA revealed a broad distinction between effects that vary as a function of how much control an agent is capable of allocating (overall performance, congruency costs, effort discounting) and effects that index how flexibly an agent can adapt to changing demands of the environment (switch costs, congruency sequence effects and proportion congruency effects). Finally, our analyses suggest that overall error rate and incongruency costs in the Stroop task best indexed the actual amount of control exerted by an agent whereas congruency sequence effect and switch costs were found to be least diagnostic.

Table 3: Regression of principal components (PC) against individual differences in EVC parameters.

Model Parameter	β	t	p
First PC, $df = 88$			
Task Automaticity	0.7867	3.58	< 0.001
Implementation Cost	-0.2582	-11.43	< 0.001
Control Capacity	0.2289	10.84	< 0.001
Reward Sensitivity	0.1870	4.53	< 0.001
Reconfiguration Cost	-0.0122	-1.61	0.111
Second PC, $df = 88$			
Task Automaticity	-0.8127	-4.10	< 0.001
Reward Sensitivity	0.0882	2.37	< 0.05
Reconfiguration Cost	-0.0586	-8.59	< 0.001
Implementation Cost	0.0435	2.14	< 0.05
Control Capacity	0.0033	0.17	0.862

Interestingly, we found that higher costs of *implementing* control were associated with lower costs of *switching* tasks. This finding is consistent with previously observed tradeoffs between cognitive stability and cognitive flexibility: higher amounts of control can reduce distractor interference but require larger reconfiguration of control signals when switching between tasks (Goschke, 2000; Musslick, Jang Jun, Shvartsman, Shenhav, & Cohen, 2018). Perhaps more surprisingly, we also found that participants with higher reconfiguration costs discounted cognitive effort *less* (i.e., were more willing to engage in demanding tasks). This finding reflects an approach-avoidance conflict inherent to demand avoidance paradigms: The more a person is engaged with a cognitively demanding task, the less they are willing to switch to an easier task (Kool, McGuire, Rosen, & Botvinick, 2010).

One limitation of the current implementation of the EVC model is its focus on 2-alternative forced choice tasks. We chose to focus on these tasks because they are amenable to analysis with the well-studied DDM (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Ratcliff, 1978). However, the DDM may be an over-simplified model for cognitive control tasks given that such tasks can involve a variety of response alternatives, as in traditional variants of the Stroop task (Stroop, 1935).

The set of relevant individual difference parameters heavily depends on the requirements of the task for cognitive control. For instance, the n-back task requires subjects to decide whether a stimulus matches the stimulus that was presented n steps before in a sequence, and has been hypothesized to involve processes of working memory updating, interference between representations held in working memory, and familiarity judgment (Chatham et al., 2011; Juvina & Taatgen, 2007). The study of individual differences in more complex tasks will require implementing more realistic process models of those tasks, such as a working memory gating model in the case of the n-back task (Chatham et al., 2011).

Altogether, these analyses suggest that individual differences in cognitive control phenomena do not necessarily re-

flect differences in someone's capacity to exert cognitive control but may as well reflect differences in task automaticity or sensitivity to reward. Accounting for differences in these variables is therefore crucial when indexing cognitive control through behavioral phenomena. However, the collinearity between simulation parameters in this analysis prevents us from teasing apart the effects of each parameter. More elaborate parameter sensitivity studies are necessary to provide more fine grained insights into the source of individual differences in cognitive control phenomena.

References

Allan, J. L., Johnston, M., & Campbell, N. (2010). Unintentional eating. what determines goal-incongruent chocolate consumption? *Appetite*, *54*(2), 422–425.

Berkman, E. T., Hutcherson, C. A., Livingston, J. L., Kahn, L. E., & Inzlicht, M. (2017). Self-control as value-based choice. *Curr Dir Psychol Sci*, *26*(5), 422–428.

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

Caplin, A., Csaba, D., Leahy, J., & Nov, O. (2018). *Rational inattention, competitive supply, and psychometrics* (Tech. Rep.). National Bureau of Economic Research.

Chatham, C. H., Herd, S. A., Brant, A. M., Hazy, T. E., Miyake, A., O'Reilly, R., & Friedman, N. P. (2011). From an executive network to executive control: a computational model of the n-back task. *Journal of cognitive neuroscience*, *23*(11), 3598–3619.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the stroop effect. *Psychological Review*, *97*(3), 332–361.

Egner, T. (2007). Congruency sequence effects and cognitive control. *Cogn Affect Behav Neurosci*, *7*(4), 380–390.

Goschke, T. (2000). Intentional reconfiguration and involuntary persistence in task set switching. *Control of cognitive processes: Attention and performance XVIII*, *18*, 331.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *J. Exp. Psychol. Gen.*, *121*(4), 480.

Juvina, I., & Taatgen, N. A. (2007). Modeling control strategies in the n-back task. In *Proceedings of the 8th international conference on cognitive modeling* (pp. 73–78).

Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitaskingan integrative review of dual-task and task-switching research. *Psychological bulletin*, *144*(6), 557.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *J. Exp. Psychol. Gen.*, *139*(4), 665.

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

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a stroop-like task. *Memory & cognition*, *7*(3), 166–174.

Manohar, S. G., Chong, T. T.-J., Apps, M. A., Batla, A., Stamelou, M., Jarman, P. R., ... Husain, M. (2015). Reward pays the cost of noise reduction in motor and cognitive control. *Current Biology*, *25*(13), 1707–1716.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1423.

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., Jang Jun, S., Shvartsman, M., Shenhav, A., & Cohen, J. D. (2018). Constraints associated with cognitive control and the stability-flexibility dilemma. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 806–811). Madison, WI.

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*. Edmonton, Can.

Navarro, D. J., & Fuss, I. G. (2009). Fast and accurate calculations for first-passage times in wiener diffusion models. *Journal of Mathematical Psychology*, *53*(4), 222–230.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*(2), 59.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *J. Exp. Psychol. Gen.*, *124*(2), 207.

Saunders, B., Milyavskaya, M., Etz, A., Randles, D., & Inzlicht, M. (2017). Reported self-control is not meaningfully associated with inhibition-related executive function: A bayesian analysis, doi: 10.1525/collabra.134.

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.

Streeter, C. C., Terhune, D. B., Whitfield, T. H., Gruber, S., Sarid-Segal, O., Silveri, M. M., ... others (2008). Performance on the stroop predicts treatment compliance in cocaine-dependent individuals. *Neuropsychopharmacology*, *33*(4), 827.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cogn Affect Behav Neurosci*, *15*(2), 395–415.