Learning to overexert cognitive control in the Stroop task

Laura A. Bustamante[1,*] **(lauraab@princeton.edu), Falk Lieder**[2,*] **(falk.lieder@berkeley.edu),**
Sebastian Musslick[1], **Amitai Shenhav**[3], **Jonathan D. Cohen**[1,4]
[1]Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA
[2]Max Planck Institute for Intelligent Systems, 72076 Tubingen, Germany
[3]Cognitive, Linguistic and Psychological Sciences and Carney Institute for Brain Sciences, Brown University, Providence, RI 02912, USA
[4]Department of Psychology, Princeton University, Princeton, NJ 08544, USA
[*] Equal Contribution

Abstract

How does the cognitive system know when and how much cognitive control to allocate to which task? According to the Learned Value of Control (LVOC) model, people learn to predict the value of cognitive control based on a linear combination of stimulus features. This model predicts that what people learn about the value of control in one situation should transfer to other situations with shared features, leading to the intriguing prediction that maltransfer can cause people to over-exert control even when it harms their performance. To test this prediction, we designed a novel color word Stroop task in which we rewarded participants differentially for exerting cognitive control (i.e. color naming) or for engaging the more automatic response (i.e. word reading) based on individual stimulus features. We test how participants learned value of control transfers to novel stimuli that share features with previously exposed stimuli and create a situation that should lead to maltransfer according to the LVOC model. Empirical data from 30 participants confirmed this prediction and supports the conclusion that maltransfer in learning about the value of control can mislead people to overexert cognitive control even when it hurts their performance.

Keywords: cognitive control, cognitive plasticity

Introduction

Cognitive control is the ability to bias the processing of stimuli in accordance with current task goals. Every day we encounter situations that require us to make appropriate decisions about the allocation of cognitive control, such as reading a paper instead of attending an incoming email notification.

Recent models have formalized control allocation as the outcome of a rational cost-benefit analysis (Shenhav, Botvinick, & Cohen, 2013; Shenhav et al., 2017; Kurzban, Duckworth, Kable, & Myers, 2013; Lieder, Shenhav, Musslick, & Griffiths, 2018). According to the expected value of control (EVC) theory the cognitive system selects the control signal that maximizes the difference between it's expected payoff and a cost based on how much control would be used (control signal intensity). Computational realizations of the EVC theory replicate interactions between motivation and cognitive control as observed in human behavior (Musslick, Shenhav, Botvinick, & Cohen, 2015), such as increases in people's willingness to allocate cognitive control when rewards are on offer (Padmala & Pessoa, 2011). An extension of this theory by Lieder et al. (2018) proposes a neurally plausible mechanism to associate stimulus features with the value of allocating control. Their Learned Value of Control (LVOC) model predicts transfer of knowledge about the value of allocating control to new situations that share stimulus features.

The forms of transfer referred to above generally lead to improved performance. However, the theory also predicts that the learning of a high value of cognitive control can lead to maltransfer from one setting to another in which allocating cognitive control turns out to be harmful, or maladaptive. For example, if people learn that it is beneficial to allocate control to the task of navigating while driving, as well as to attending to a mobile device to encode a text message, then without sufficient experience to the contrary they might reasonably (though regrettably) infer that it would be advantageous to allocate control to both tasks in a setting where both are available. Here, we test the predictions of the LVOC model experimentally, and manipulate the association of stimuli with the benefits of allocating control, to generate circumstances in which the acquisition of such associations can lead to maltransfer. We focus on conditions in which such associations favor the maladaptive allocation of control over an automatic alternative that would have resulted in the better performance.

LVOC Model

According to the LVOC model, people learn to predict the EVC of specifying the control signal c in situation s, $\text{EVC}(s,c)$, from the stimulus features associated with each control signal, $\mathbf{f}(c)$;

$$\text{EVC}(s,c) \approx \text{LVOC}(f,c;w) = \sum_i w_i^{(f\times c)} \cdot f_i \cdot c - \text{cost}(c), \quad (1)$$

where w_i is the weight of the i^{th} feature and $\text{cost}(c)$ is the cost of exerting the control signal. To illustrate this, consider the Stroop task (Stroop, 1935), in which participants are shown a series of color words (e.g. **RED**, **GREEN**, **BLUE**, etc.), and asked either to read the word (WR) or name the color (CN) in which it is displayed (e.g. respond to the incongruent stimulus **RED** by saying "green"). To represent Stroop stimuli with the colors red and green and the words **RED** and **GREEN**, the features would include $f_1(c) = \text{colorIsRed} \cdot c, f_2(c) = \text{colorIsGreen} \cdot c, f_3(c) = \text{wordIsRed} \cdot c, f_4(c) = \text{wordIsGreen} \cdot c$ as well as the overall (global) value of exerting cognitive control in the context of the experiment $f_5(c) = c$. Feature weights

$\mathbf{w} = (w_1, \cdots, w_5)$ are learned by Bayesian linear regression of the experienced value of control $R - \text{cost}(c)$ onto the features $\mathbf{f}$, where R is the reward experienced upon control allocation. We applied this model to predict people's accuracy and reaction times in the novel Stroop experiment described below. Following Musslick et al. (2015), our model translates control signals into reaction times and error rates via a drift-diffusion model with the drift rate:

$$d = c^{\star} \cdot y_{\text{color}} \cdot d_{\text{controlled}} + (1 - c^{\star}) \cdot y_{\text{word}} \cdot d_{\text{automatic}}, \quad (2)$$

where $y_{\text{color}}, y_{\text{word}} \in \{-1, 1\}$ are the responses associated with the color or word respectively, and $d_{\text{controlled}}$ and $d_{\text{automatic}}$ are the drift rates of the automatic (WR) process and the controlled (CN) process.

Experiment

We studied the maltransfer of cognitive control in a variant of the Stroop task (Stroop, 1935) in which participants were exclusively presented with incongruent stimuli for which the color and word lead to different responses. It is assumed that naming the color of an incongruent Stroop stimulus requires the allocation of control to overcome interference from an automatic word reading process, leading to worse performance for CN compared to WR (Cohen, Dunbar, & McClelland, 1990).

We divided the experiment into two phases: an initial Mapping Phase in which participants learned the value of a controlled response (CN) associated with individual stimulus features (colors and words); and a second Transfer Phase, in which participants were tested on novel combinations of previously encountered stimulus features (see Figure 1). We specifically examined how participant's learned value of allocating cognitive control in the Mapping Phase would affect their decision to allocate cognitive control (CN) in the Transfer Phase. We hypothesized that participants learn a high value of exerting control (CN) for stimuli with features that predict a reward for a CN response in the Mapping Phase. For instance, rewarding the CN response for stimuli that are composed of color green or the word **RED** should lead participants to learn a high value of exerting control (CN) whenever either feature is present. Critically, when both features are combined to a new stimulus (**RED**) in the Transfer Phase participants' should be biased toward exerting cognitive control, even though they have never encountered this stimulus. We used eight colors and eight words. Half of each (yellow/green/red/blue,**YELLOW**/**GREEN**/**BLUE**/**BLUE**) were of experimental interest, and the other half (white/orange/brown/pink, **WHITE**/**ORANGE**/**BROWN**/**PINK**) were used as controls.

Mapping Phase. Participants were instructed[1] that they could either WR or CN. In two parts of the Mapping Phase participants learned the response associations between the features of interest and the CN or WR task. For example, in Mapping Phase Part 1 every time the color yellow was presented the correct response was WR (other colors of interest were also mapped). In Mapping Phase Part 2, every time the word **YELLOW** was presented the correct response was CN (other words of interest were also mapped). A display indicated whether they received or missed the reward of five points for responding correctly. Each of the mapping phases consisted of 160 trials.

Transfer Phase. The transfer phase consisted of six trial types, each trial type is defined based on the response assigned to them in Transfer Phase relative to response assigned to their features in the Mapping Phase (see Figure 1). We examined participant's propensity to allocate cognitive control to new stimuli (combinations of words and colors) in the Transfer Phase of the experiment. We composed new stimuli out of features (colors and words) which had been individually associated with a positive value of control (e.g, the color green and the word **RED**) but were never shown in conjunction in the Mapping Phase of the experiment. Critically, WR was now the more highly rewarded response for those stimuli. Note that the feature-response mapping resembles that of an exclusive OR (XOR) operation: Each of the two features was associated with color-naming. However, their conjunction would now require participants to engage the automatic response to read the word. We referred to these trials as *consistently-mapped neither* trials and based on the LVOC model predict maltransfer (increased errors and response times) for these trials.

The five other trial types are as follows. First, we considered consistently-mapped color trials for which two features of interest are combined. CN is the correct response in the Transfer Phase, and the color feature was mapped to CN in the Mapping Phase. The word feature of consistently-mapped color trials was mapped to WR, nonetheless CN is the correct response in the Transfer Phase and participants must learn this contingency. Similarly, for consistently-mapped word trials, two features of interest were combined and CN is the correct response. The word feature was mapped to CN in the Mapping Phase where the color feature was mapped to the WR response. These trials correspond to the 'OR' rule in the XOR problem; when one feature that is associated with CN is present, the correct response is to CN. For consistently-mapped both trials, both features were associated with WR in the Mapping Phase and WR was the correct response in the Transfer Phase. We parametrically manipulated the frequency of consistently-mapped color and consistently-mapped word (collectively, *consistently-mapped color/word*) trials and expected increased experience with *consistently-mapped color/word* trials would increase maltransfer to *consistently-mapped neither* trials. In a between subjects design we included 0%, 20%, and 50% *consistently-mapped color/word* trial frequencies in the Transfer Phase. We also included a set of control trial types in the Transfer Phase to measure non-feature-based learning and to ensure CN was rewarded equally often as WR. We refer to them as control trial types because they do not share features

[1] Participants given given no prior information about which stimuli features rewarded CN versus WR.

with, and should not be affected by the *consistently-mapped color/word* trial frequency manipulation. We used novel combinations of the control colors and words for the control trial types. Control colors and words were used in the Mapping Phases in combination with colors and words of interest, and were equally often associated with the CN and WR response during Mapping. For *WR control* trials, we combined a subset of the control colors and control words and the correct response was to WR. For *CN control* trials we combined an different subset of the control colors and control words and the correct response was to CN. *WR control* trials were frequency matched to *consistently-mapped neither* trials and served as a measure of non-feature-based learning of the value of control as *consistently-mapped color/word* trial frequency increased. *CN control* trials were balanced against the frequency of *consistently-mapped color/word* trials so that CN was the correct response for half of all trials in the transfer phase in all frequency groups (i.e., when *consistently-mapped color/word* trial frequency was 20%, *CN control* trial frequency was 30%). We tested 30 participants (10 per frequency condition) at Princeton University. Participants received bonus compensation for points they earned.

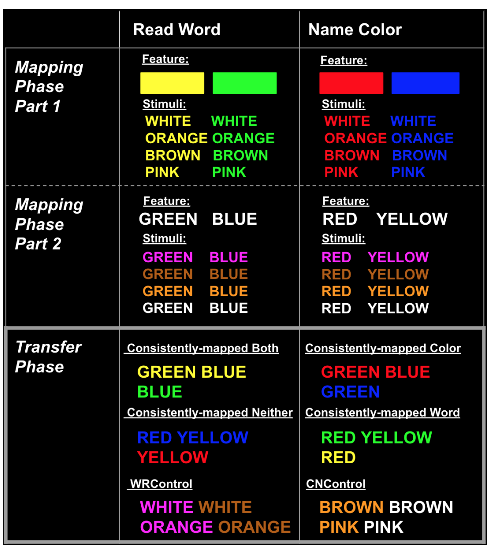

Figure 1: Experiment Stimuli. Columns: which response is rewarded. Rows: experiment phase.

Model fit to behavior

To directly compare the control signal output by the LVOC model and human performance we performed a hierarchical DDM fit to participants' accuracy and reaction times in the Transfer Phase (HDDM version 0.6.0 in Python 3.4; Wiecki, Sofer, & Frank, 2013). We were particularly interested in the rate of evidence accumulation, or drift rate, as an indicator for the strength of processing of color or word relative to which response was correct on each trial type. We fit the LVOC model to the group-level estimates of the participants' drift rates for each trial type in each experimental condition. Free parameters in the model were the prior distribution of the weights, and the drift rates for the color-naming and word-reading process (averaged across all participants). We found that the best fitting parameters were $(1.32, 3.22, -0.17, 0.11)$. Using these best fitting parameters we simulated a 30 participant experiment 100 times. Our dependent measure was the mean drift rate applied on each trial type in the simulations.

Results

Behavior

Consistent with our predictions, there was a significant increase in errors on *consistently-mapped neither* trials (CN responses) with increasing *consistently-mapped color/word* trial frequency ($\beta = 0.03276$, $p < 1.17e-05$). HDDM fits indicated a significant decrease in drift rate towards the correct response (WR), and near reversal toward the incorrect response, with increasing frequency of *consistently-mapped color/word*, indicating decreased strength of processing of the WR response (see Figure 2A; sample mean frequency 0% = 0.824 ± 0.130, frequency 20% = 0.316 ± 0.126, frequency 50% = -0.099 ± 0.126). We found no change in reaction time across *consistently-mapped color/word* frequency groups. The fits did indicate a lower threshold for the 50% frequency group compared to the 0% and 20% groups, suggesting that shift in the speed-accuracy contributed at least in part to the higher error rates in this group.

The only trial type for which human performance was not aligned with the LVOC model predictions was *WR control* trials. These trials, that appeared only in the Transfer Phase, did not share features with *consistently-mapped color/word* trials, and were associated with WR as the correct response. If learning was feature-specific, we would not expect performance on *WR control* trials to be influenced by the frequency of *consistently-mapped color/word* trials. Contrary to this prediction, errors on *WR control* trials increased with an increase in the frequency of *consistently-mapped color/word* trials ($\beta = 0.03465$, $p < 4.47e-05$). Correspondingly, drift rate towards the correct response (WR) decreased as *consistently-mapped color/word* trial frequency increased (see Figure 2A; sample mean frequency 0% $= 0.96 \pm 0.13$, frequency 20% = 0.59 ± 0.12, frequency 50% = 0.22 ± 0.13). Linear mixed effects modeling did not reveal any significant change in reaction time across frequency groups.

Model fit and model comparison

We found that the model matched the drift rates estimated directly from the data (see Figure 2) with the exception of *WR control* trials. For those trials, the model did not capture the decrease in drift rate with increased frequency of *consistently-mapped color/word* trials. This was likely because the model is driven largely by feature-based learning. To determine that the complexity of the model was justified by its fit to the data, we performed quantitative model comparisons against an alternative Win-Stay Lose-Shift (WSLS) model that switched between CN and WR following errors and repeated its choice following correct response. Our analyses indicated that the LVOC model explained the data significantly better than the simpler alternative model (BIC LVOC=4.5 vs. BIC WSLS=707.8)

One possible explanation of the deterioration of people's performance on *WR control* trials with the frequency of *consistently-mapped color/word* trials could be a difference in how the global value of control is learned as compared to the stimulus-specific weights. *WR control* performance would suggest this learned global weight is larger in higher frequency groups. Consistent with the idea that, with sparse data, people learn the EVC for more general (i.e., shared) characteristics of a task situation, and only with more training do they develop dedicated (i.e., distinct or separated) representations of the EVC for more specific features of the circumstance, we allowed the global weight to be learned more rapidly than other weights. We did so by increasing the prior precision of the global weight which allows faster learning because there is less uncertainty. Importantly, this mechanism is agnostic to the direction in which the global weight should change. We set the precision of the model's prior on the weights of feature and feature conjunctions (64.9) to be significantly less precise than the prior on the global weight for control signal intensity (21.13). We found that this manipulation allowed the model to capture the pattern of performance in the *WR control* trials, as well as all of the other effects previously exhibited.

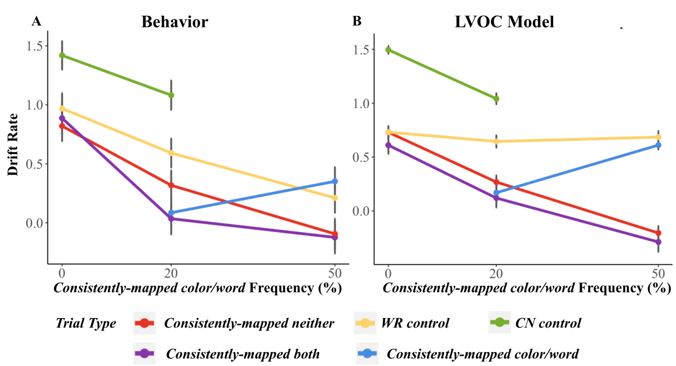

Figure 2: A) Behavior fit drift rate. Drift rates are towards the rewarded (correct) response for a trial type. B) Fit of LVOC model to the drift rates of human subjects. LVC model captures qualitative effects in drift rates for all trial types except *WR control* trials (yellow).

Discussion

We designed a novel Stroop task to test the hypothesis that the value of control is learned and that this learning will occur based on features present in the environment when the outcome of allocating control is realized. Participants and the LVOC model showed a significant propensity to choose the control-demanding response (CN) for stimuli combining features that were previously associated with higher reward for the controlled response in the Mapping Phase, even when there was greater reward for selecting the less demanding response (WR) for that combination of features (*consistently-mapped neither* trials) in the Transfer Phase. These findings strongly suggest that people are attentive to and learn about the value of allocating control, and generalize what they learn to new circumstances. At the same time, the findings suggest that people use relatively simple learning rules (in this case, one that appears to be a linearly additive) that, at least in this experiment, did not show sensitivity to non-linear associations (i.e., the XOR rule used to determine the rewarded response in the Transfer Phase). We also included trials that probed non-feature based transfer of the value of control; surprisingly, these also yielded evidence for the over-allocation of control. Post-hoc modeling suggested that over-generalization of the value of control may reflect a bias towards learning lower dimensional representations of the value of control (i.e., the predictive value of a situation irrespective of its features) early in learning. Lastly, our results suggests that learning about the value of control may be subject to maltransfer in situations in which there is a non-linear relationship between the value of control predicted for features present and the actual value of control for the conjunction of those features; for example, the value of driving and of attending to a mobile device.

Acknowledgments This work was made possible by the John Templeton Foundation, grant number 57876, and the National Institutes of Health, grant number T32MH065214.

References

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.

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*(6), 661–679.

Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018, 04). Rational metareasoning and the plasticity of cognitive control. *PLOS Computational Biology*, *14*(4), 1-27.

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 *2nd multidiscip. conf. reinf. learn. decis. mak.*

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., Cohen, J., & Botvinick, M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*.

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

Wiecki, T., Sofer, I., & Frank, M. (2013). Hddm: Hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics*, *7*.