Reward masks the learning of cognitive control demand

Bettina Bustos[1,2], Jiefeng Jiang[1,2] and Wouter Kool[3]

[1]Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA 52242
[2]Cognitive Control Collaborative, University of Iowa, Iowa City, IA 52242
[3]Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, MO 63130

Corresponding authors:

Wouter Kool, Ph. D
wkool@wustl.edu
314-935-8729
Washington University
CB 1125
One Brookings Drive
St. Louis, MO 63130

And

Jiefeng Jiang, Ph. D
jiefeng-jiang@uiowa.edu
319-467-4565
PBSB 373
340 Iowa Avenue
Iowa City, IA 52242

Abstract

Cognitive control refers to a set of cognitive functions that modulate other cognitive processes to align with internal goals. Recent research has shown that cognitive control can flexibly adapt to internal and external factors such as reward, effort, and environmental demands. This suggests that learning processes track changes in these factors and drive an optimization process to determine how cognitive control should be applied in changing situations. In real life, multiple factors often simultaneously affect how cognitive control is deployed. However, previous studies mainly concern how cognitive control adjusts to changes in a single factor. Here, we investigate how cognitive control learns to adjust to two concurrently changing factors: statistical regularity in cognitive control demand and performance-contingent reward. We consider two competing hypotheses: reward promotes cognitive control to adjust to cognitive control demand, and the processing of reward information obstructs the adaptation to cognitive control demand. In our experiment, statistical regularity in cognitive control demand is manipulated within subjects such that some stimuli require higher levels of cognitive control than others. Reward is manipulated across subjects. Using a computational model that captures temporal changes in cognitive control, we find that in the absence of reward, participants can adjust to different levels of cognitive control demand. Importantly, when performance-contingent reward is available, participants fail to adapt to changes in cognitive control demand. The findings support the hypothesis that reward blocks the learning of cognitive control.

Introduction

Our ability for cognitive control enables us to reconfigure mental functions to implement effortful, non-routine, and goal-directed behavior (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Egner, 2017; Miller & Cohen, 2001). A key feature of cognitive control is its flexibility to adapt to changing environmental and internal factors (Jiang, Beck, Heller, & Egner, 2015; Jiang, Heller, & Egner, 2014; Monsell, 2003). Such flexibility is thought to optimize the tradeoff (Ritz, Leng, & Shenhav, 2022; Shenhav, Botvinick, & Cohen, 2013) between potential gains, such as reward (Botvinick & Braver, 2015), goal achievement, opportunity costs (Otto & Daw, 2019), and potential costs such as cognitive effort (Kool & Botvinick, 2018; Kool, McGuire, Rosen, & Botvinick, 2010; Shenhav et al., 2017; Westbrook, Kester, & Braver, 2013).

These factors are typically studied in isolation. For example, humans adjust cognitive control in the response to demands posed by the environment. In the classic Stroop task (Stroop, 1935), they increase their use of cognitive control to the proportion of demanding incongruent trials (e.g., the word GREEN printed in red; Braem et al., 2019; Gratton, Coles, & Donchin, 1992; Jacoby, Lindsay, & Hessels, 2003). Such adaptation does not only happen across temporal contexts (Egner, 2007; Schmidt, 2013; Spinelli, Morton, & Lupker, 2022; Ullsperger, Bylsma, & Botvinick, 2005) but also in response to statistical contingencies between items and demand (Bugg & Dey, 2018; Bugg, Jacoby, & Chanani, 2011; Chiu, Jiang, & Egner, 2017; Jiang, Bramao, et al., 2020). Computationally, this flexibility can be captured by the reinforcement-learning framework (Jiang et al., 2015; Jiang et al., 2014), which posits that the brain learns to predict future demand based on a running average of experienced (in)congruency, and adjusts cognitive control accordingly.

Cognitive control also flexibly adapts in response to available reward. Indeed, performance-contingent reward manipulations increase accuracy and reduce response times through the enhancement of cognitive control (Balleine & Dickinson, 1998; Botvinick & Braver, 2015; Chaillou, Giersch, Hoonakker, Capa, & Bonnefond, 2017; Chiew & Braver, 2014; Fröber & Dreisbach, 2014). In real life, cognitive control needs to flexibly adapt to a host of factors, including reward and predictable statistical structures, simultaneously. For example, rewards may be given in a challenging task that affords the learning of cognitive control demand. However, little is known about how these factors jointly affect flexible cognitive control.

We consider two hypotheses of how cognitive control simultaneously adapts to changes in performance-related incentives and statistical contingencies predicting the demand for cognitive control. On the one hand, performance-contingent rewards may increase the motivation to tailor the attentional state to the structure of cognitive control demand. On the other hand, the increased motivation to engage cognitive control triggered by reward may interfere with the learning of updated demand contingencies. For example, because both reward and learning of statistical contingency produce prediction errors (PEs, the discrepancy between predicted and actual outcome), these may interfere and lead to reduced learning of statistical contingency in the presence of reward. Moreover, structure learning has been characterized as carrying a control cost (Collins, 2017), thus increased cognitive control on a focal task may reduce availability for control applied to updating statistical structure.

To adjudicate between these hypotheses, we manipulated the statistical structure of cognitive control demands by changing the item-specific proportion congruence (ISPC) in a picture-word Stroop task. In an initial 'Inducer' phase of the experiment, certain stimuli mostly appeared in congruent trials (i.e., having low cognitive control demand), whereas others mostly

appeared in incongruent trials (i.e., having high cognitive control demand). Humans are sensitive to these differences between trial types, adopting a more focused attentional state for items that are mostly incongruent and a more relaxed one for items that are mostly congruent (Bugg & Crump, 2012; Bugg & Dey, 2018; Bugg et al., 2011; Crump, Gong, & Wmiken, 2006; Logan & Zbrodoff, 1979; Suh & Bugg, 2021). In a second, Diagnostic, phase this statistical structure was altered, with all items appearing with 50% percent congruence. If participants are sensitive to this change in the statistical structure of the task, item-specific cognitive control settings should regress from their previously focused and relaxed states to identical and intermediate states for all stimuli. To test whether reward and structure learning interact, one group of participants received performance-contingent reward in the Diagnostic phase, whereas the other group did not. We analyzed the data using conventional analyses of error rates and reaction time (RT) data and computational modeling using a reinforcement-learning framework (Chiu et al., 2017; Sutton & Barto, 2018). To preview the results, we find that the presence of performance-contingent reward reduces the learning of the ISPC, thus supporting the hypothesis that reward interferes with the learning of statistical contingency of cognitive control. To our knowledge, this work provides the first evidence of the interaction between reward and the learning of cognitive control demand in driving adaptive behavior.

Method

Participants

No-reward condition. One hundred and twenty-nine participants from Washington University in St. Louis provided informed consent and earned class credit for participation. Twenty-one participants were removed for not meeting the accuracy threshold of 80%, resulting

in a final sample of 108 participants (Age M = 19.4, SD = 1.09, 83 female, 23 male, 2 preferred not to answer). All procedures were approved by Washington University in St. Louis Institutional Review Board.

Reward condition. One hundred and twenty-two participants were recruited on Amazon Mechanical Turk to participate in the experiment and provided informed consent. Six participants were removed for not meeting the accuracy threshold of 80%, resulting in a final sample of 116 participants (Age M = 40.66, SD = 11.85, 49 female, 67 male). All procedures were approved by Washington University in St. Louis Institutional Review Board.

Stimuli

The stimuli used in both conditions of this study were a subset of a larger set of stimuli from Bugg et al. (2011). The stimulus set consisted of sixteen line drawings of four birds, four cats, four fish, and four dogs, along with four capitalized words (BIRD, CAT, FISH, DOG). On each trial, a word (45px) was superimposed on a line drawing (215 x 161px). Both the word and the line drawing were presented at the center of the screen.

Procedure

Participants completed either a Reward condition or a No-reward condition. All procedures were held constant across both conditions in the Inducer phase of the experiment with no reward. The only difference between the two conditions occurred in the Diagnostic phase, namely that in the Reward condition participants were given the opportunity to receive performance-contingent monetary reward, while participants in the No-reward condition did not receive reward**.**

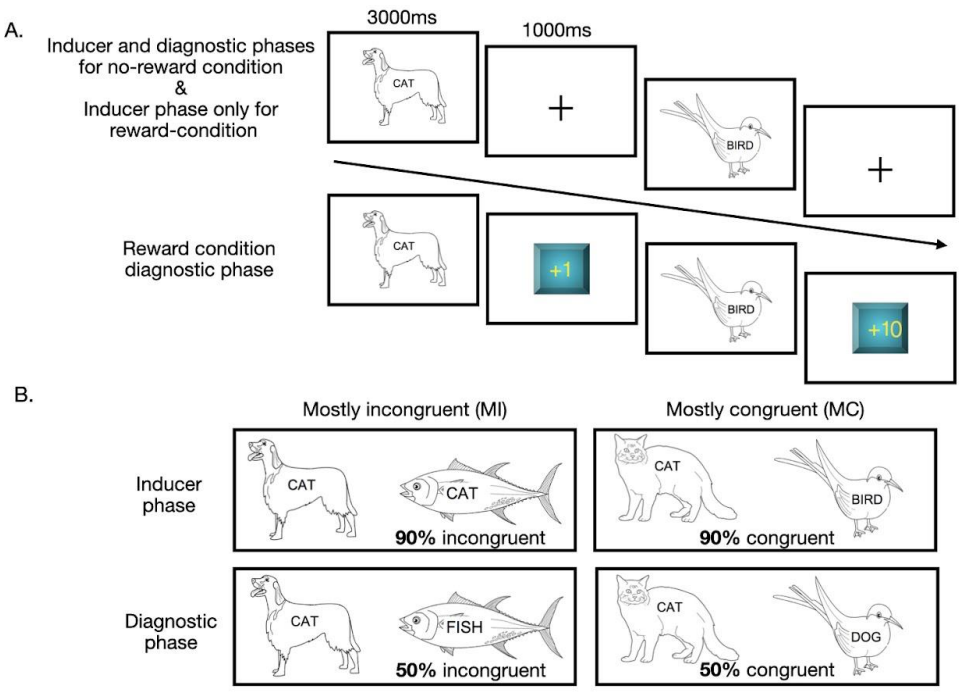

Figure 1. *Experimental design.* (A) Trial sequence. Exemplar trials for Diagnostic and Inducer phases for each condition. (B) ISPC for each phase. ISPC is the same in both conditions. MI and MC categories are counterbalanced across participants.

Each trial started with the presentation of a picture-word Stroop stimulus until a valid response was made or until the response deadline of 3000ms was met (Fig. 1A). Participants were instructed to respond to the animal that appeared in the picture-word Stroop stimuli using previously learned response keys. They were also instructed to ignore the animal names superimposed on the picture. On congruent trials, the identity of the picture and the word matched (e.g., a picture of a cat with the word CAT superimposed), whereas on incongruent trials, the identity of the picture and the word conflicted (e.g., a picture of a cat with the word DOG superimposed). Stimulus-response mappings (e.g., the correspondence between keys on the keyboard [A, S, D, or F] and animal pictures [dog, fish, cat, bird]) were randomized across

participants. Trials were separated by an inter-trial interval (ITI) of 1000ms. A crosshair was presented at the center of the screen during the ITI. The crosshair was replaced was replaced with a reward stimulus on trials in which participants received reward (only in the Diagnostic phase of the Reward condition).

After the instructions, participants first completed the Inducer phase, which consisted of three blocks of 240 trials each for a total of 720 trials. Specifically, trials with two types of animal pictures were 90% congruent (mostly congruent, or MC condition, Fig. 1B), and trials with the other two types of animal pictures were congruent 10% of the time (mostly incongruent, or MI condition, Fig. 1B). The animal-congruence mappings were randomized across participants.

After the Inducer phase, participants in both conditions completed a Diagnostic phase consisting of one block of 192 trials. In this phase, participants performed the same task, except that the previously learned ISPC biases were changed – all stimuli had the same 50% probability of appearing in a congruent trial. Note that in the Diagnostic phase, the congruence condition (MC or MI) of a stimulus refers to its ISPC in the Inducer phase. Critically, participants in the Reward condition were instructed that quick and accurate responses would be probabilistically rewarded (Fig. 1A). Specifically, one MI animal category and one MC animal category carried a 90% chance of paying out high reward (10 points) for accurate responses while the other two animal categories had a 10% chance of paying out low reward (1 point). On trials in which they received reward, participants were presented with either “+1” or “+10” points displayed within a hexagonal teal-colored shape during the ITI inter-trial interval (1000 ms). In the No-reward condition, the procedure of the Diagnostic phase remained identical to the Reward condition, with the exception that no reward was provided.

Before starting the main task, participants went through a practice phase, in which they were extensively instructed on (1) the correct stimulus-response mappings, and (2) the picture-word Stroop task. Participants practiced the main task for 16 trials. To prevent the practice phase from confounding the learning in the main task, the practice phase used different line drawings (two per category) than those used in the main task. The practice phase is explained in detail in the Supplemental Materials.

Conventional statistical analysis

RTs faster than 200ms or slower than 3,000ms were excluded, consistent with prior research with this task (e.g., Bugg & Dey, 2018; Bugg et al., 2011). In addition, only correct responses were included in the analysis of RTs. Both exclusions eliminated less than 1% of the trials. Mean RTs and error rates for each phase and each condition are presented in Table 1. We conducted 2 x 2 repeated-measures analyses of variance (rmANOVAs) with factors of ISPC (MC and MI) and Trial Type (Congruent and Incongruent) for both error rates and RT in the Induction phase as well as the Diagnostic phase. Additionally, to examine the dynamics of ISPC learning in the Diagnostic phase, we split the Diagnostic phase into two halves and repeated the aforementioned rmANOVAs on each half. Finally, we examined whether the amount of reward obtained in the Diagnostic phase of the Reward condition had any effect on ISPC using independent sample *t*-tests. Here, the ISPC effect was quantified as the interaction between PC and Trial Type using the difference in congruency effect (i.e., incongruent - congruent) between MC and MI conditions (Bugg et al., 2011). All statistical analyses were conducted using Pingouin (version number: 0.3.12; Vallat (2018)).

Model-based analysis

Modeling the learning of cognitive control demand. We used a reinforcement learning model to capture the dynamics of how participants tracked the statistical structure of cognitive control demand (i.e., ISPC in this study) throughout the different phases. The model predicts the probability of encountering an incongruent trial for each animal category across the task. A model was applied to each category to account for the manipulation of PC at the animal category level. The predicted incongruency of the presented category was modeled and updated on each trial as $P_i \leftarrow P_i + \alpha(c - P_i)$, where P_i is the predicted incongruency (i.e., the predicted probability of encountering an incongruent trial) for animal category *i* presented on the current trial, and c encodes the incongruency at the current trial, where congruent and incongruent trials were coded as 0 and 1, respectively. In other words, on each trial, the updated P_i (i.e., on the left side of the equation) is a sum of the current value of P_i (i.e., on the right side of the equation) and an updating term based on the current-trial PE, defined as $(c - P_i)$, weighted by a learning rate α. For the Inducer phase, all P_is were initialized as 0.5 to reflect a neutral prediction of incongruency. The initial P_i of the Diagnostic phase was set to the final P_i of the Inducer phase to simulate the retention of ISPC. The model incorporates two learning rates (αs), one for the Inducer phase and one for the Diagnostic phase.

Behavioral model fitting. The learning rates are the only free parameters of the model. To model how P (subscripts are removed to indicate a vector of all trial-wise values for the variable) influences RT, a linear model using and $|PE|$ as a predictor was constructed to predict trial-wise RT (Chiu et al., 2017; Jiang, Bramao, et al., 2020; Muhle-Karbe, Jiang, & Egner, 2018). The absolute value of PE is used because only the magnitude but not the direction of PE is required to examine the strength of learning. Specifically, a larger $|PE|$ indicates more deviation of the predicted probability of incongruency from the actual (in)congruency, and when used to guide

cognitive control, will cause suboptimal processing and slower RT (Jiang et al., 2015; Jiang et al., 2014). In addition to the model-based regressors, we included five binary predictor variables. The first coded for the congruency of the current trial, c, to account for the congruency effect, and the four resulting binary predictor variables coded for each of the four animal categories present on the current trial to account for potential biases that may differentially influence RTs for each of the animal categories. All trials were included in the trial-by-trial modeling of P and $|PE|$, but only correct trials were included in the multiple linear regression predicting RT.

To determine the best-fitting αs, we implemented a standard minimization function (using the L-BFGS-B algorithm) that was run with 30 different randomly selected starting points for each participant to avoid local minima. For each participant, the parameter fits of the model with the smallest sum of squared errors (SSE) of predicted RT over all trials was selected. Within each phase, the learning rate was shared by all category-specific learning models. The optimization procedure also produced the regression coefficient of |PE| (β_{PE}), which indicates how much RT scales with the magnitude of prediction error.

*Model-based statistical analysis***.** Because β_{PE} was generated based on the best-fitting learning model, it quantifies the strength of learning: As discussed above, a positive β_{PE} indicates that RT slows with PE magnitude and thus suggests the behavior is consistent with the learning model (Jiang et al., 2015). In other words, a positive coefficient suggests that the participant learned the current ISPC and reacted to deviations from it by slowing their response. To probe the difference in ISPC learning between conditions, we conducted an independent-samples *t*-test on β_{PE} between conditions (Reward or No-reward) in the Inducer and Diagnostic phases separately. The same analysis was also performed on learning rates. Welsh's *t*-test was used to account for unequal sample sizes between the Reward and No-reward conditions.

Results

For conventional statistical analysis, we report the descriptive statistics of error rates and RTs across trial types, phases and conditions in Table 1.Their statistical results are listed in Table 2 (RTs) and 3 (error rates). The full results are in Supplementary Notes 1 and 2. ANOVA results of the No-reward condition are originally reported in Bustos, Colvett, Bugg, and Kool (in preparation). Briefly, the error rates data showed two patterns: (1) higher error rates in incongruent than congruent trials and (2) no significant difference in ISPC effects between the reward conditions, which indicates that the difference in ISPC effects in RT data were unlikely to be attributable to speed-accuracy tradeoff. RT data from the Inducer phase showed that participants in both conditions learned the ISPC (Fig. 2). RT data further suggested reduced learning of the ISPC in the Reward condition compared to the No-reward condition in the second half of the Diagnostic phase. We also found no difference in the ISPC effect between the low- and high-reward items in the Diagnostic phase of the Reward condition (Supplementary Note 3) and hence ruled out the possibility of reward amount being a confounding factor of the results. In summary, reward seems to interfere with masks learning of the ISPC. In the below, we provide support for this hypothesis using model-based measures that better capture learning dynamics.

Mean Reaction Time (ms) and Error Rates

Phase	Condition	Item type PC	DV	Trial type		Congruency effect
				Con	Inc	
Induction	No-reward	MC	RT	750 (117)	860 (159)	110 (81)
			Error rate	4.4% (0.03)	7.3% (0.06)	2.9% (0.06)

		MI	RT	753 (129)	806 (140)	53 (60)
			Error rate	4.1% (0.04)	5.9% (0.04)	1.7% (0.04)
	Reward	MC	RT	801 (171)	914 (202)	113 (84)
			Error rate	4.1% (0.04)	6.2% (0.07)	2.1% (0.06)
		MI	RT	816 (183)	860 (186)	44 (60)
			Error rate	4.3% (0.05)	6.0% (0.05)	1.5% (0.05)
Diagnostic	No-reward	MC	RT	770 (160)	836 (177)	66 (76)
			Error rate	5.5% (0.05)	7.5% (0.07)	2.0% (0.05)
		MI	RT	771 (166)	813 (170)	42 (73)
			Error rate	5.4% (0.06)	6.6% (0.6)	1.2% (0.05)
	Reward	MC	RT	749 (159)	835 (189)	86 (75)
			Error rate	4.5% (0.07)	6.7% (0.08)	2.2% (0.06)
		MI	RT	756 (159)	817 (171)	61 (62)
			Error rate	4.1% (0.05)	5.8% (0.06)	1.7% (0.05)

Table 1. Note. Values in parentheses indicate standard deviation. PC = proportion congruence; RT = reaction time; MC = mostly congruent; MI = mostly incongruent; DV = dependent variable. Note that items in the Diagnostic phase were 50% congruent.

Conventional statistical results: RT

Phase	Condition	Effect	DOF	F value	P value	Effect size η_p^2
Inducer	No-reward	Main effect of trial type	(1,107)	258.53	<.001	.71
	Reward		(1,115)	191.86	<.001	.63
	No-reward	Main effect of ISPC	(1,107)	18.85	<.001	.15
	Reward		(1,115)	8.79	<.001	.07
	No-reward	Trial type × ISPC interaction	(1,107)	36.63	<.001	.26
	Reward		(1,115)	87.80	<.001	.43
Diagnostic	No-reward	Main effect of trial type	(1,107)	114.37	<.001	.52
	Reward		(1,115)	174.84	<.001	.61
	No-reward	Main effect of ISPC	(1,107)	12.15	.146	.02
	Reward		(1,115)	.86	.36	.01
	No-reward	Trial type × ISPC interaction	(1,107)	5.57	.02	.05
	Reward		(1,115)	15.75	<.001	.12
Diagnostic: First half	No-reward	Main effect of trial type	(1,107)	71.05	<.001	.40
	Reward		(1,115)	126.51	<.001	.52
	No-reward		(1,107)	.35	.56	.003

	Reward	Main effect of ISPC	(1,115)	.84	.36	.01
	No-reward	Trial type × ISPC interaction	(1,107)	10.54	.002	.09
	Reward		(1,115)	8.56	.004	.07
Diagnostic: Second half	No-reward	Main effect of trial type	(1,107)	54.21	<.001	.34
	Reward		(1,115)	133.8	<.001	.54
	No-reward	Main effect of ISPC	(1,107)	3.05	.08	.03
	Reward		(1,115)	.49	.48	.004
	No-reward	Trial type × ISPC interaction	(1,107)	.03	.86	<.001
	Reward		(1,115)	8.51	.004	.07

Table 2. Note. Items in the Diagnostic phase were 50% congruent. DOF = degree of freedom.

Conventional statistical results: Error rate

Phase	Condition	Effect	DOF	F value	P value	Effect size η_p^2
Inducer	No-reward	Main effect of trial type	(1,107)	38.61	<.001	.26
	Reward		(1,115)	22.28	<.001	.16
	No-reward	Main effect of ISPC	(1,107)	6.14	.015	.05
	Reward		(1,115)	.11	.74	<.001
	No-reward	Trial type × ISPC interaction	(1,107)	4.26	.04	.04
	Reward		(1,115)	.99	.32	.01
Diagnostic	No-reward	Main effect of trial type	(1,107)	21.02	<.001	.16
	Reward		(1,115)	28.34	<.001	.20
	No-reward	Main effect of ISPC	(1,107)	2.1	.15	.02
	Reward		(1,115)	1.71	.74	.01
	No-reward	Trial type × ISPC interaction	(1,107)	1.53	.22	.01
	Reward		(1,115)	.54	.46	.004
Diagnostic: First half	No-reward	Main effect of trial type	(1,107)	16.61	<.001	.13
	Reward		(1,115)	10.70	.001	.09
	No-reward	Main effect of ISPC	(1,107)	.35	.55	.003
	Reward		(1,115)	.72	.40	.001
	No-reward	Trial type × ISPC interaction	(1,107)	2.13	.15	.02
	Reward		(1,115)	3.15	.08	.03
Diagnostic: Second half	No-reward	Main effect of trial type	(1,107)	5.14	.03	.06
	Reward		(1,115)	28.35	<.001	.20
	No-reward	Main effect of ISPC	(1,107)	2.08	.15	.02
	Reward		(1,115)	1.71	.74	.01
	No-reward	Trial type × ISPC interaction	(1,107)	.05	.83	.02
	Reward		(1,115)	.54	.46	.004

Table 2. Note. Items in the Diagnostic phase were 50% congruent. DOF = degree of freedom.

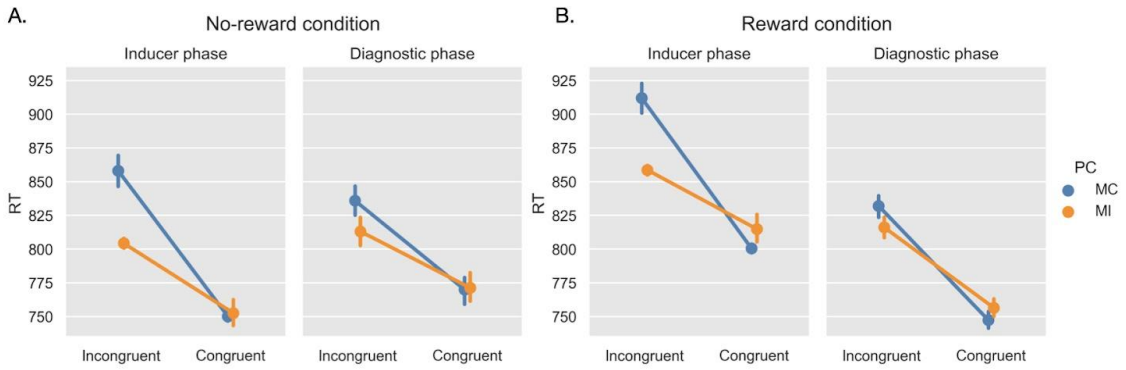

Figure 2. *RT results.* (A) RTs by trial type (congruent or incongruent) and proportion congruency (mostly congruent, MC, or mostly incongruent, MI) in the No-reward condition for each phase. (B) RTs by trial type (congruent or incongruent) and proportion congruency (mostly congruent, MC, or mostly incongruent, MI) in the Reward condition for each phase.

Model-based results

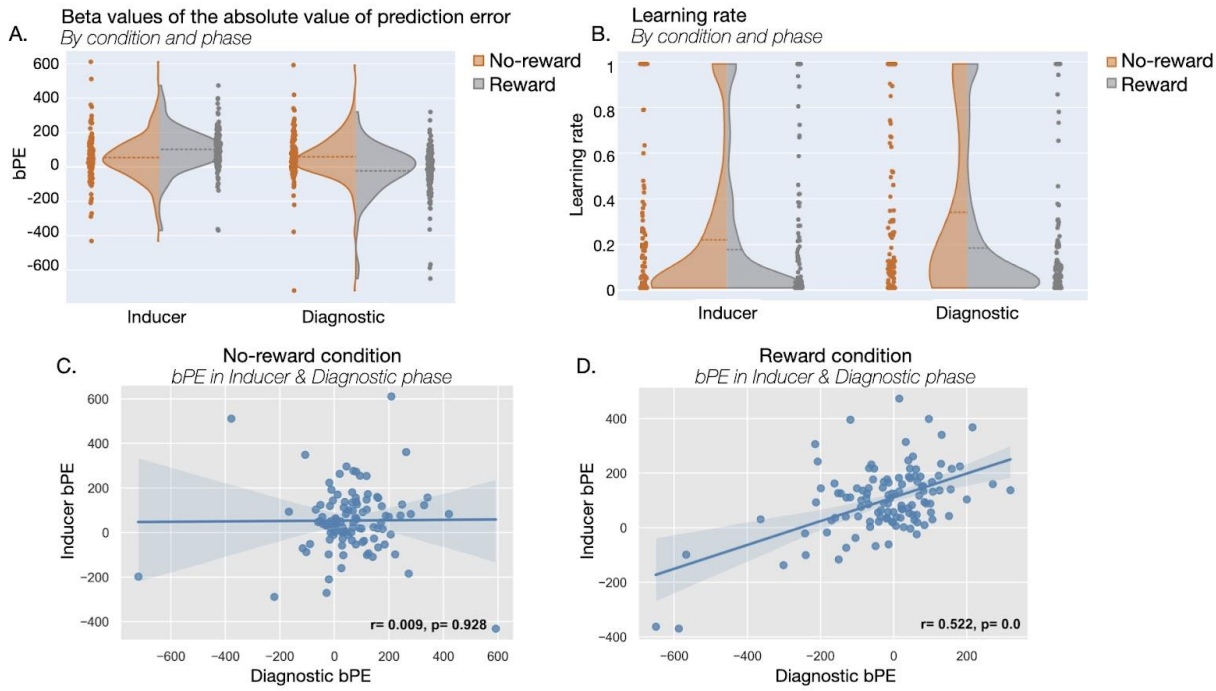

Figure 3. *Model-based RT results.* **(A)** Regression weights of the absolute value or prediction error for each condition and each phase. (B) Fit learning rates for each condition and phase. (C) Correlation between the regression weight of |PE| in the Inducer phase and in the Diagnostic phase for the No-reward condition. (D) Correlation between the regression weight of |PE| in the Inducer phase and in the Diagnostic phase for the Reward condition.

Inducer phase. We used trial-wise model estimates of unsigned PE of incongruency to predict RT in a linear model. Previous studies have shown that trial-wise RT scales with PE in an ISPC design (Chiu et al., 2017; Jiang, Bramao, et al., 2020). Replicating these findings, we found that β_{PE} was significantly above zero in both conditions (No-reward condition: $t(107) = 3.50$, $p < .001$, $d = .34$; Reward condition: $t(115) = 10.7$, $p < .001$, $d = .99$, Fig. 3A). As a validation, we then tested the correlation between the regression weight of PE and ISPC effect in RT. In both conditions, the behavioral ISPC effect was highly correlated with the regression weight of $|PE|$ (No-reward: $r = 0.59$, $p < .001$, Reward: $r = .63$, $p < .001$). Importantly, the average regression weight of $|PE|$ (i.e., β_{PE}) was significantly smaller in the No-reward condition than the Reward condition, $t(211) = -2.63$, $p = .01$, $d = .35$ (No-reward: $M = 54$, SE = 13.90, Reward: $M = 101$, SE = 11.45), suggesting that the Reward group learned ISPC better than the No-reward group. To take into consideration this difference, the β_{PE} of the Inducer phase was used as a baseline in the Diagnostic phase analysis (see below). In addition to β_{PE}, we also compared learning rates between the two conditions. To account for the non-Gaussian distributions of learning rates a non-parametric test, the Mann–Whitney *U* test (Wilcoxon rank-sum test) was implemented to compare learning rates within the Inducer phase and revealed no significant difference between

conditions (U = 6711, p =.*34*, Common-Language Effect Size (CLES) = .54; No-reward: M = 0.22, SE = .03, Reward: M = 0.17, SE = 0.03, Fig. 3B).

Diagnostic phase. Similar to the Inducer phase, we used β_{PE} as a measure of ISPC learning. In other words, as the learning models adapt to the new ISPC, a strong modulation of PE on RT (i.e., large β_{PE}) indicates that participants learned the (now unbiased) ISPC and used it to influence cognitive control and behavior. Conversely, a weak effect of β_{PE} indicates no learning of ISPC. Thus, by comparing this measure between conditions, we tested whether the presence of reward affects the learning of the new ISPC in the Diagnostic phase. An independent t-test showed that the β_{PE} in the No-reward condition (M = 59, SE = 14.0) was larger compared to the Reward condition (M = -23, SE = 13.72; $t(221) = 4.19$, $p < .001$, Fig. 3A). This finding suggests that the Reward condition showed reduced learning of the new (unbiased) ISPC compared to the No-Reward condition. This is consistent with the split-phase RT analysis reported above, in which participants in the Reward condition were shown to carry the outdated ISPC from the Inducer phase throughout the second half of the Diagnostic phase, unlike participants in the No-reward condition.

One alternative explanation to this finding is that the subjects in the Reward condition were in general worse learners of ISPCs. To rule out this possibility, we used the Inducer phase as a baseline and tested the change of β_{PE} from the Inducer to the Diagnostic phase and found that β_{PE} change was also significantly larger in the Reward condition (M = 83, SE = 17.95) than the No-reward condition (M = -.26, SE = .02, $t(213) = -4.39$, $p < .001$, d = .58). This result indicates that the Reward condition reduced their learning of ISPC compared to the No-reward condition. An additional alternative hypothesis is that the subjects in the Reward condition are more sensitive to the biased ISPC in the Inducer phase (i.e., 90% and 10%) compared to the

unbiased ISPC in the Diagnostic phase (i.e., 50%). As β_{PE} scales with the strength of ISPC learning, this hypothesis would predict that subjects showing larger β_{PE} in the Inducer phase will show smaller β_{PE} in the Diagnostic phase. However, opposite to this prediction, we observed a significant positive cross-subject relationship of the PE weights between the two phases in the Reward condition (r = 0.52, p < .001, Fig. 3C). As a control, when no reward was presented, this relationship was absent in the No-reward condition between phases (r = .009, p = .93, Fig. 3D). This difference in correlations was statistically significant (z = 4.19, p < .001). Thus, these findings did not support the alternative explanation.

The Mann–Whitney U test comparing learning rates within the Diagnostic phase revealed a significant difference between conditions (U = 7802, p = .001, CLES = .62; Reward condition: M = .18, SE = 0.03; No-reward condition: M = .34, SE = 0.04, Fig. 3B). The learning rate increased significantly in the No-reward condition from the Inducer phase to the Diagnostic phase $t(214) = 2.53$, p = .01, d = 0.34, but did not differ significantly between phases for the Reward condition, $t(230) = -.18$, p = 0.85, d =.02, Finally, we found that the change in learning rate between phases was significant between conditions, $t(195) = 1.98$, p = .049, d = .28 (Fig. 3b). The significant increase in the learning rate from the Inducer to the Diagnostic phase for the No-reward study further supports the notion that learning of the ISPC in the Diagnostic phase was faster in the No-reward than the Reward condition. In summary, we observed slower learning of the ISPC in the Diagnostic phase in the Reward than in the No-reward condition, manifested in both PE modulation on RT and learning rates.

Discussion

Extant work probing the influence of reward on cognitive control has examined its contribution to performance in several domains including attention (Botvinick & Braver, 2015; Chiew & Braver, 2014; Engelmann & Pessoa, 2014; Notebaert & Braem, 2015; Yee & Braver, 2018) and memory encoding (Adcock, Thangavel, Whitfield-Gabrieli, Knutson, & Gabrieli, 2006; Miendlarzewska, Bavelier, & Schwartz, 2016; Murty, LaBar, & Adcock, 2016; Spaniol, Schain, & Bowen, 2014; Wittmann et al., 2005). In the present study we aimed to study how cognitive control adapts to multiple external factors, such as reward and statistical regularity. Using a between-subject design that manipulates the presence of performance-contingent reward in a variant of the Stroop task, our results first show that, in the Inducer phase, when no reward was available, participants in both group learned associations between stimulus features (here, the animals in the pictures) to guide cognitive control, such that larger Stroop effects (suggesting less cognitive control) were observed for items that had been MC than MI items. This finding replicates the classic ISPC effect (Bugg et al., 2011; Jacoby et al., 2003). Importantly, in the following Diagnostic phase, where all items were now unbiased, participants in the Reward condition showed slower learning of the new ISPC than the No-reward condition. Specifically, our model-based analyses revealed that unlike participants in the Reward condition, participants in the No-reward condition elicited a significant increase in β_{PE} from the Inducer phase to the Diagnostic phase further demonstrating that participants in the No-reward condition were more sensitive to the changes in their environment. Together, these results provide evidence for the interference of reward incentives on the updating of learned control strategies.

There are several non-exclusive accounts for why reward reduces the learning of associations between items and cognitive control demand. First, in the Reward condition, participants are simultaneously learning the structure of reward (e.g., the RT threshold to obtain

reward, and which items yield high reward, etc.) and the ISPC. Both forms of learning result in prediction error in the striatum and the midbrain (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Kim, Sul, Huh, Lee, & Jung, 2009; Schönberg, Daw, Joel, & O'Doherty, 2007; Schultz, Dayan, & Montague, 1997) and thus may cause competition and subsequent masking of ISPC PE (Chiu & Egner, 2019; Chiu et al., 2017) by reward PE. Second, because learning in the Reward condition involves two factors (reward and ISPC), this may carry an additional cognitive cost and make the learning of ISPC less effective (Collins, 2017). Third, participants may prioritize reward learning and direct more resources (e.g., attention) to it (Niv, 2019). It has been shown that attention magnifies PE in perception (Jiang, Summerfield, & Egner, 2013). Similarly, the learning of ISPC, when not attended, may receive only a weak learning signal of PE, leading to reduced learning. Last but not least, both reward and ISPC learning occur at the item level, which may cause memory interference.

Our findings may seem to be inconsistent with previous research demonstrating that performance-contingent reward increases proactive cognitive control (Chiew & Braver, 2014; Fröber & Dreisbach, 2014). We argue that both the previous and current findings reflect a similar strategy under the framework of the expected value of control (EVC) theory (Shenhav et al., 2013). The EVC theory posits that cognitive control is applied to balance reward and cost (e.g., cognitive effort). When no reward is available, ISPC can be learned such that high cognitive control is applied only when necessary (e.g., when encountering MI items) to reduce effort. When reward is contingent on performance, the balance of cognitive control is biased towards reward, allowing for more effort to be applied to both MC and MI items. As a result, it is no longer necessary to distinguish between MC and MI items. This is crucial in the current Reward condition, as reward depends on fast response. Thus, performance-contingent reward will lead to

two compatible consequences corresponding to the previous and the current findings. First, cognitive control can be applied proactively (i.e., prior to the onset of the item). Second, the re-learning of ISPC becomes less valuable, as there is no need to separate MC and MI items.

ISPC is considered to reflect reactive cognitive control (i.e., congruence is only available after the item is displayed). This leads to the intriguing possibility that this reward-congruence learning tradeoff can be different in the case of proactive cognitive control. For example, in a context-specific proportion congruency task (Jiang, Wang, Guo, Fernandez, & Wagner, 2020; King, Korb, & Egner, 2012), the expected congruence of the current trial (i.e., the context) can be proactively cued without revealing the correct response. If this paradigm were combined with the current design, we would predict the opposite pattern of results. Specifically, if the present findings were due to cost-benefit tradeoff in cognitive control, proactive cognitive control, which can be deployed before the trial starts, can benefit more from the latest congruence information than reactive cognitive control. Therefore, in a rewarded diagnostic phase, participants would be able to improve performance through proactive control, and therefore we would predict increased learning in this condition. Indeed, Braem, Hickey, Duthoo, and Notebaert (2014) found that reward promoted increased sensitivity to context-specific congruency effects, lending credibility to this hypothesis. Though, in these studies, participants were not afforded the opportunity to re-learn previously established statistical regularities in control demands. Alternatively, reduced learning in the rewarded condition (a replication of the current results) would support the hypothesis that reward and congruence PE are encoded by a shared mechanism, and that the reduced learning is driven by their interference.

More broadly, this study made an initial attempt to gauge the relative strength of different factors affecting cognitive control. This is important for understanding cognitive control in real

life, in which cognitive control needs to simultaneously adapt to multiple factors such as reward, cognitive control demand, effort, context and motivation. Although the EVC theory has the potential to explain the adaptation of cognitive control, it relies on the understanding of how benefits and costs of multiple factors interact. The interaction seems to be complex, as the factors may not be treated independently. For example, our result shows that reward masks ISPC learning. They indicate that a comprehensive view of how the metacontrol of cognitive control needs to take into account interaction between multiple factors.

Acknowledgement

J.J. is supported by the National Institute of Mental Health (R01MH131559).

Open Practices Statement

References

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. (2006). Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron, 50*(3), 507-517.

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology, 37*(4-5), 407-419.

Botvinick, M., & Braver, T. (2015). Motivation and cognitive control: from behavior to neural mechanism. *Annu Rev Psychol, 66*, 83-113.

Botvinick, M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychol Rev, 108*(3), 624-652.

Braem, S., Bugg, J. M., Schmidt, J. R., Crump, M. J., Weissman, D. H., Notebaert, W., et al. (2019). Measuring adaptive control in conflict tasks. *Trends in Cognitive Sciences, 23*(9), 769-783.

Braem, S., Hickey, C., Duthoo, W., & Notebaert, W. (2014). Reward determines the context-sensitivity of cognitive control. *Journal of Experimental Psychology: Human perception and performance, 40*(5), 1769.

Bugg, J. M., & Crump, M. J. (2012). In support of a distinction between voluntary and stimulus-driven control: A review of the literature on proportion congruent effects. *Frontiers in psychology, 3*, 367.

Bugg, J. M., & Dey, A. (2018). When stimulus-driven control settings compete: On the dominance of categories as cues for control. *Journal of Experimental Psychology: Human Perception and Performance, 44*(12), 1905.

Bugg, J. M., Jacoby, L. L., & Chanani, S. (2011). Why It Is too Early to Lose Control in Accounts of Item-Specific Proportion Congruency Effects. *Journal of Experimental Psychology-Human Perception and Performance, 37*(3), 844-859.

Bustos, B., Colvett, J., Bugg, J., & Kool, W. (in preparation). Humans do not avoid reactive control demands.

Chaillou, A.-C., Giersch, A., Hoonakker, M., Capa, R. L., & Bonnefond, A. (2017). Differentiating motivational from affective influence of performance-contingent reward on cognitive control: the wanting component enhances both proactive and reactive control. *Biological psychology, 125*, 146-153.

Chiew, K. S., & Braver, T. S. (2014). Dissociable influences of reward motivation and positive emotion on cognitive control. *Cognitive, Affective, & Behavioral Neuroscience, 14*(2), 509-529.

Chiu, Y. C., & Egner, T. (2019). Cortical and subcortical contributions to context-control learning. *Neurosci Biobehav Rev, 99*, 33-41.

Chiu, Y. C., Jiang, J., & Egner, T. (2017). The Caudate Nucleus Mediates Learning of Stimulus-Control State Associations. *J Neurosci, 37*(4), 1028-1038.

Collins, A. G. (2017). The cost of structure learning. *Journal of cognitive neuroscience, 29*(10), 1646-1655.

Crump, M. J. C., Gong, Z. Y., & Wmiken, B. (2006). The context-specific proportion congruent Stroop effect: Location as a contextual cue. *Psychonomic Bulletin & Review, 13*(2), 316-321.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69*(6), 1204-1215.

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

Egner, T. (2017). *The Wiley handbook of cognitive control*. Southern Gate, Chichester, West Sussex, UK: Wiley Blackwell.

Engelmann, J. B., & Pessoa, L. (2014). Motivation sharpens exogenous spatial attention.

Fröber, K., & Dreisbach, G. (2014). The differential influences of positive affect, random reward, and performance-contingent reward on cognitive control. *Cognitive, Affective, & Behavioral Neuroscience, 14*(2), 530-547.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology: General, 121*(4), 480.

Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: stroop process dissociations. *Psychon Bull Rev, 10*(3), 638-644.

Jiang, J., Beck, J., Heller, K., & Egner, T. (2015). An insula-frontostriatal network mediates flexible cognitive control by adaptively predicting changing control demands. *Nat Commun, 6*, 8165.

Jiang, J., Bramao, I., Khazenzon, A., Wang, S. F., Johansson, M., & Wagner, A. D. (2020). Temporal Dynamics of Memory-guided Cognitive Control and Generalization of Control via Overlapping Associative Memories. *J Neurosci*.

Jiang, J., Heller, K., & Egner, T. (2014). Bayesian modeling of flexible cognitive control. *Neurosci Biobehav Rev*.

Jiang, J., Summerfield, C., & Egner, T. (2013). Attention sharpens the distinction between expected and unexpected percepts in the visual brain. *J Neurosci, 33*(47), 18438-18447.

Jiang, J., Wang, S.-F., Guo, W., Fernandez, C., & Wagner, A. D. (2020). Prefrontal reinstatement of contextual task demand is predicted by separable hippocampal patterns. *Nature communications, 11*(1), 1-12.

Kim, H., Sul, J. H., Huh, N., Lee, D., & Jung, M. W. (2009). Role of striatum in updating values of chosen actions. *Journal of neuroscience, 29*(47), 14701-14712.

King, J. A., Korb, F. M., & Egner, T. (2012). Priming of control: implicit contextual cuing of top-down attentional set. *J Neurosci, 32*(24), 8192-8200.
Kool, W., & Botvinick, M. (2018). Mental labour. *Nature human behaviour, 2*(12), 899-908.
Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *J Exp Psychol Gen, 139*(4), 665-682.
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 and Cognition, 7*, 166-174.
Miendlarzewska, E. A., Bavelier, D., & Schwartz, S. (2016). Influence of reward motivation on human declarative memory. *Neuroscience & Biobehavioral Reviews, 61*, 156-176.
Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu Rev Neurosci, 24*, 167-202.
Monsell, S. (2003). Task switching. *Trends Cogn Sci, 7*(3), 134-140.
Muhle-Karbe, P. S., Jiang, J., & Egner, T. (2018). Causal Evidence for Learning-Dependent Frontal Lobe Contributions to Cognitive Control. *J Neurosci, 38*(4), 962-973.
Murty, V. P., LaBar, K. S., & Adcock, R. A. (2016). Distinct medial temporal networks encode surprise during motivation by reward versus punishment. *Neurobiology of learning and memory, 134*, 55-64.
Niv, Y. (2019). Learning task-state representations. *Nature neuroscience, 22*(10), 1544-1553.
Notebaert, W., & Braem, S. (2015). Parsing the effects of reward on cognitive control. In *Motivation and cognitive control* (pp. 117-134): Routledge.
Otto, A. R., & Daw, N. D. (2019). The opportunity cost of time modulates cognitive effort. *Neuropsychologia, 123*, 92-105.
Ritz, H., Leng, X., & Shenhav, A. (2022). Cognitive control as a multivariate optimization problem. *Journal of Cognitive Neuroscience, 34*(4), 569-591.
Schmidt, J. R. (2013). Questioning conflict adaptation: proportion congruent and Gratton effects reconsidered. *Psychon Bull Rev, 20*(4), 615-630.
Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *Journal of Neuroscience, 27*(47), 12860-12867.
Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593-1599.
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., et al. (2017). Toward a Rational and Mechanistic Account of Mental Effort. *Annu Rev Neurosci*.
Spaniol, J., Schain, C., & Bowen, H. J. (2014). Reward-enhanced memory in younger and older adults. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 69*(5), 730-740.
Spinelli, G., Morton, J. B., & Lupker, S. J. (2022). Both task-irrelevant and task-relevant information trigger reactive conflict adaptation in the item-specific proportion-congruent paradigm. *Psychonomic Bulletin & Review*, 1-13.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology, 18*(6), 643.
Suh, J., & Bugg, J. M. (2021). On the automaticity of reactive item-specific control as evidenced by its efficiency under load. *Journal of Experimental Psychology: Human Perception and Performance, 47*(7), 908.
Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning : an introduction* (Second edition. ed.). Cambridge, Massachusetts: The MIT Press.
Ullsperger, M., Bylsma, L. M., & Botvinick, M. M. (2005). The conflict adaptation effect: It's not just priming. *Cognitive, Affective, & Behavioral Neuroscience, 5*(4), 467-472.

Vallat, R. (2018). Pingouin: statistics in Python. *J. Open Source Softw., 3*(31), 1026.
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.
Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H.-J., & Düzel, E. (2005). Reward-related FMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron, 45*(3), 459-467.
Yee, D. M., & Braver, T. S. (2018). Interactions of motivation and cognitive control. *Current opinion in behavioral sciences, 19*, 83-90.