Learning when effort matters: Neural dynamics underlying updating and adaptation to changes in performance efficacy

Ivan Grahek[1†*], Romy Frömer[1†], & Amitai Shenhav[1]

1. *Department of Cognitive, Linguistic, & Psychological Sciences and Carney Institute for Brain Science, Brown University, Providence, RI 02912, USA*

[†] These authors have contributed equally to the study.

* Corresponding author. E-mail address: ivan_grahek@brown.edu

Abstract

Cognitive control is critical for performing well on most tasks, but sometimes success isn't determined by whether we perform well. To know how much effort to invest into a task, we must know how much our performance matters in our current environment. Here, we examined how people generate and update expectations about the *efficacy* of their performance, and how they adjust their control allocation accordingly. Participants performed an incentivized cognitive control task, and over time we varied the extent to which rewards were determined by performance or at random. We found that participants updated their efficacy estimates based on deviations from current expectations, and that model-based estimates of such *efficacy prediction errors* were encoded by an EEG signature of updating (P3b). Participants adjusted their effort based on their updated efficacy estimates (investing more effort with higher expected efficacy), reflected in behavioral and neural signatures of control allocation on subsequent trials.

Keywords: Motivation; Cognitive control; Learning; Performance efficacy; Reward rate; EEG

Cognitive control is critical for achieving most goals, but it is effortful (Botvinick & Cohen, 2014; Shenhav et al., 2017). Determining whether and how much control to invest in a task requires a person to weigh these effort costs against the potential benefits of control (Kool & Botvinick, 2018; Manohar et al., 2015; Verguts et al., 2015). These potential benefits not only reflect the amount of reward that can be expected for performing well on a task but also, and critically, the extent to which performance matters for obtaining those rewards (the *efficacy* of performance) (Frömer et al., 2021; Shenhav et al., in press; Shenhav et al., 2013). For instance, a student may be more motivated to invest effort into writing an essay for a competition if they believe that this essay will play an important role in determining their success than if they believe that the competition will be primarily determined by performance-unrelated factors (e.g., social status, idiosyncratic criteria). The amount of cognitive effort we invest in a task is therefore substantially determined by how much efficacy we expect to have with that task. Such efficacy expectations must be derived from past experiences that indicated how worthwhile our efforts were in a given environment. Understanding how people come to learn when their efforts are efficacious is critically important for understanding why they come to be more or less motivated in certain environments, which is particularly relevant for understanding mental disorders (Grahek et al., 2019; Moscarello & Hartley, 2017). However, in spite of its importance, relatively little is known about the mechanisms by which people learn how efficacious their cognitive efforts are.

Past work has provided critical insights into related aspects of how people adapt cognitive control based on learned incentives. For instance, a substantial body of research has examined how people learn about the amount and frequency of reward in their environment (Daw & Tobler, 2013; Dolan & Dayan, 2013; Niv et al., 2006), including the extent to which

learning differs depending on the valence of feedback (e.g., whether an outcome was better or worse than expected; Chambon et al., 2020; Collins & Frank, 2014; Garrett & Daw, 2020; Lefebvre et al., 2017). While this work has traditionally investigated the relationship between learned value and action selection, recent work has shown that people also adapt their control states based on learned value (Lieder et al., 2018). People allocate more or less cognitive control to a task depending on the reward they have come to expect based on recent trials (Bustamante et al., 2021; Otto & Daw, 2019). Similarly, people adapt their control based on the expected difficulty of a given task (for a review see: Chiu & Egner, 2019). Behavioral and electrophysiological evidence shows that people allocate more control when, based on previous learning, they expect a task to be more difficult (Bejjani et al., 2018; Bugg et al., 2011; Jiang et al., 2020; Nigbur et al., 2015). Research into these difficulty-related control adaptations provides insight into how people learn about and adapt to changes in the amount of cognitive control required to achieve a given level of performance. However, in the real world, achieving that level of performance may not always have bearing on whether a person will be rewarded (Shenhav et al., in press). Research on cognitive control adaptation has yet to examine how people learn and adapt as the performance efficacy of their environment changes.

We recently showed how the expected efficacy of mental effort can be experimentally dissociated from expectation of reward and task difficulty (Frömer et al., 2021). Participants performed a cognitive control task (Stroop) and we varied the amount of reward they could obtain on a given trial as well as whether the rewards were determined by their performance (high efficacy) or whether the rewards were determined at random (low efficacy). We showed that people allocate more control when they expect to have high compared to low efficacy, reflected in higher amplitudes of an EEG index of proactive control (the contingent negative

variation [CNV]) and in improved performance. However, unlike in most real-world examples, in this study participants were explicitly cued with which level of efficacy to expect prior to each trial, the efficacy associated with a cue was fixed through the session, and efficacy levels were always at one of two extremes. Therefore, there are two key open questions which are unaddressed to date: how do people learn about their efficacy in a given task, and how do they update these efficacy estimates?

In this study, we use computational modeling and single-trial EEG analyses to study the mechanisms through which efficacy estimates are learned, updated, and used to guide control allocation. We show that efficacy estimates can be learned from feedback through iterative updating based on weighted prediction errors (Sutton & Barto, 2018). Using model-based analyses of a putative neural marker of learning (P3b; Fischer & Ullsperger, 2013; Nassar et al., 2019), we show that efficacy prediction errors and corresponding learning rates can be used to predict how efficacy estimates are adjusted following relevant changes in the environment. We further show that efficacy estimates learned in this way are used to guide the allocation of control. Extending our previous findings (Frömer et al., 2021), we also show that a marker of proactive control allocation (the CNV; Grent-’t-Jong & Woldorff, 2007; Schevernels et al., 2014) increases with increasing model-based efficacy estimates. Taken together, these results show that, like other motivationally relevant variables used to guide the allocation of cognitive control, efficacy estimates are learned and updated based on feedback, leveraging general cognitive and neural mechanisms of predictive inference.

Results

To investigate how efficacy estimates are learned, and how they affect control allocation, we recorded EEG while 40 participants performed a modified version of the Stroop task (Figure

1A). Across trials we varied whether reward outcomes ($0.50 vs. $0.00) were determined by a participant's performance on a given trial (responding accurately and below a pre-determined response time criterion; *performance-based* trials) or whether those outcomes were determined independent of performance (based on a weighted coin flip; *random* trials). Over the course of the session, we gradually varied the likelihood that a given trial would be performance-based or random such that, at some points in the experiment, most trials were performance-based (high efficacy level), and at other points most trials had random outcomes (low efficacy level) (Figure 1B). Importantly, unlike in our previous study (Frömer et al., 2021), participants were not told whether a given trial would be performance-based (maximal efficacy) or random (minimal efficacy), but instead had to estimate their current efficacy level based on recent trial feedback, which indicated both the reward outcome ($0.50 vs. $0.00) and how that outcome was determined (performance-based [button symbol] vs. random [dice symbol]). We held reward rate constant across both feedback types by yoking reward rate on random trials to the reward rate on performance-based trials, and counter-balanced the time-course of the gradual change in efficacy (see Methods for details). To capture changes in efficacy expectations over the course of the session, we probed participants every 4-8 trials (averaging 44.3 probes per participant) to report their current estimates of efficacy. These efficacy probes were interleaved with probes asking participants to estimate the current reward rate, serving as foils and for control analyses.

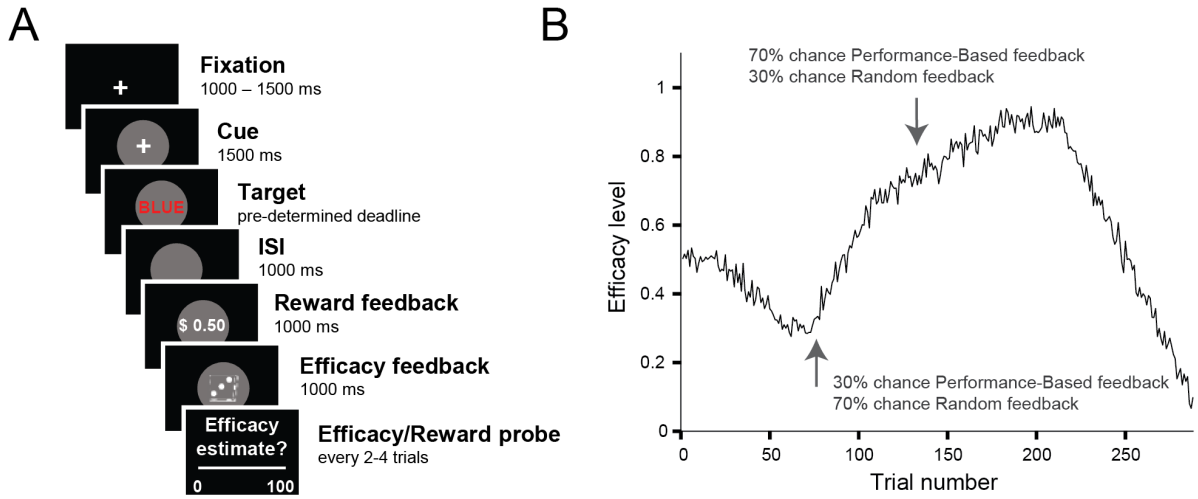

Figure 1. Manipulating expected efficacy and assessing learning. A. Trial Schematic. On each trial participants saw a cue (gray circle), predicting the onset of a Stroop stimulus (target), and were then sequentially presented with reward and efficacy feedback. On half of the trials, efficacy feedback was presented first, and on the other half reward feedback was presented first. Every 2-4 trials participants were subsequently asked to estimate their current efficacy level or reward rate. **B. Efficacy manipulation.** We let the probability of performance-based vs. random feedback continuously drift over the course of the experiment (inversed for one half of the sample). Arrows mark time points with low and high efficacy, respectively. When efficacy was low, rewards were more likely to be random, whereas when efficacy was high, rewards were more likely to be performance-based.

Participants dynamically update efficacy expectations based on feedback

To determine whether and how participants learned their current efficacy levels, we first analyzed the influence of previous efficacy feedback (whether outcomes on a given previous trial had been performance-based or random) on one's current subjective estimates of efficacy. We found very strong evidence that participants adjusted their subjective efficacy upward or downward depending on whether the most recent trial was Performance-Based or Random (b = 0.14; 95% CrI [0.13, 0.16]; $p_{b<0} = 0$; $BF_{10} > 100$), and that this remained true (but to diminishing degrees) up to five trials back (all $p_{b<0} < 0.01$; Figure 2A; Table S1). These results suggest that participants were dynamically updating their efficacy estimates based on efficacy feedback.

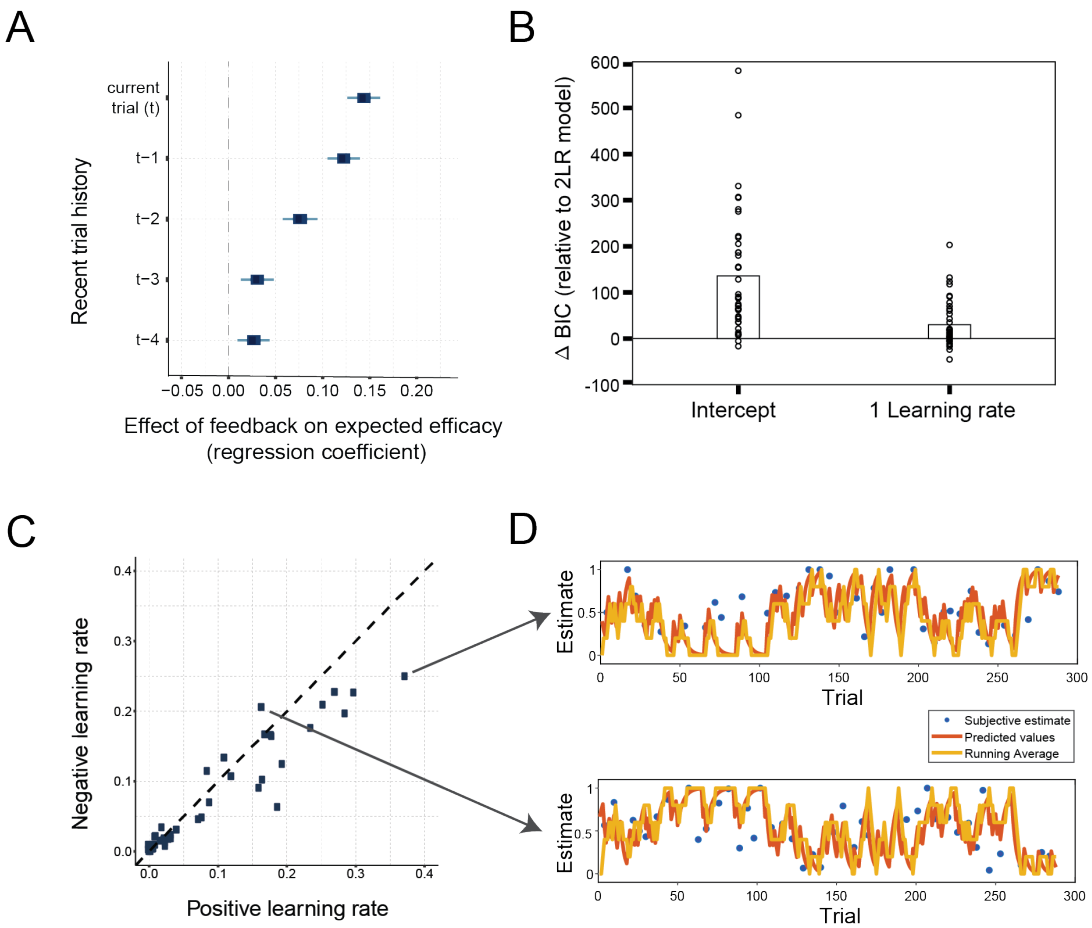

Figure 2. Efficacy learning is captured by a reinforcement learning model separate learning rates for performance-based and random feedback. A. Efficacy estimates track recent efficacy feedback. Regression weights for the influence of the current and previous contingent vs. random feedback on the subjective efficacy estimate. Error bars represent 50% and 95% highest density intervals. **B. A separate learning rate model captures efficacy learning best.** BICs of fitted intercept only and one learning rate model relative to two learning rate models are plotted for each participant and favor the two learning rate model. **C. Learning rate biases vary between participants.** Positive and negative learning rate estimates are plotted for all participants. Points below the diagonal indicate higher learning rates for performance-based compared to random feedback, and points above the opposite. **D. Example learning trajectories.** Subjective and model-based efficacy estimates, and a running average of the previous 5 efficacy feedbacks, for a participant with a higher learning rate for the contingent compared to random feedback (upper) and a participant with the reverse bias (lower).

This pattern of learning was accounted for by a standard reinforcement learning (RL) algorithm, the delta-rule, according to which efficacy estimates are updated in proportion to the prediction error between the expected efficacy and the efficacy feedback on a given trial (i.e., whether a given outcome was determined by performance or not) (Sutton & Barto, 2018).

Interestingly, consistent with studies of reward learning (Chambon et al., 2020; Collins & Frank, 2014; Garrett & Daw, 2020; Lefebvre et al., 2017; Niv et al., 2012), we found that the RL model that best accounted for our data allowed efficacy estimates to be updated differently from trials that were more efficacious than expected (Performance-Based feedback) than from trials that were less efficacious than expected (Random feedback), by having separate *learning rates* scaling prediction errors in the two cases. Even when accounting for model complexity, this Two Learning Rate model outperformed a One Learning Rate model as well as a baseline model that fits a single constant efficacy estimate and no learning (Intercept Model) (Figure 2B). We found that the two learning rates for this best-fit model varied across the group (Figure 2E), but did not find that one was consistently higher than the other (b = 0.02; 95% CrI [-0.02, 0.05]; $p_{b<0}$ = 0.203; BF_{01}=16.41), despite most participants (80%) tending to learn more from Performance-Based than Random trials. Taken together, these results show that participants dynamically updated their expected efficacy based on trial-by-trial feedback, and that they did so differentially based on whether the trial was more or less efficacious than expected. The fitted models further enable us to generate trial-by-trial estimates of expected efficacy and efficacy prediction errors, which we use in model-based analyses of behavior and neural activity below. Note that in all the following analyses we control for reward rate estimates obtained from models fit to reward feedback (for details see below).

The feedback-related P3b indexes updating of efficacy expectations

To investigate the neural mechanism underlying feedback-based learning of efficacy, we probed the centroparietal P3b ERP component (Figure 3A), an established neural correlate of prediction-error based learning (Fischer & Ullsperger, 2013; Nassar et al., 2019). If the P3b

indexes feedback-based updating of *efficacy* predictions, as it does for reward predictions, we would expect this ERP to demonstrate several key markers of such a learning signal. First, we would expect the P3b to reflect the extent to which efficacy feedback (Performance-Based vs. Random) deviates from the current level of expected efficacy. In other words, the P3b should track the magnitude of the *unsigned efficacy prediction error (PE)* on a given trial. We tested this by examining how the amplitude of the P3b to a given type of efficacy feedback varied with model-based estimates of the participant's efficacy expectation on that trial. If the P3b signaled efficacy PE then its amplitude should scale inversely with the expected probability of a given type of feedback (i.e., how *unexpected* that feedback is), thus correlating *negatively* with expected efficacy on trials providing Performance-Based feedback and correlating *positively* with expected efficacy on trials providing Random feedback. In addition to overall higher P3b to performance-based compared to random feedback (b = 0.90; 95% CrI [0.55, 1.24]; $p_{b<0}$ = 0; BF_{10} > 100; Figure 2B), we found exactly this predicted interaction between feedback type and expected efficacy (b = -2.77; 95% CrI [-4.08, -1.48]; $p_{b>0}$ = 0; BF_{10} > 100; Figure 2B; Table S2), with the P3b amplitude decreasing with model-based estimates of expected efficacy on Performance-Based trials (b = -1.93; 95% CrI [-3.07, -0.75]; $p_{b>0}$ = 0; BF_{10} = 6.04) and increasing numerically, but not robustly with expected efficacy on Random trials (b = 0.84; 95% CrI [-0.28, 2.00]; $p_{b<0}$ = 0.11; BF_{10} = 0.43). Accordingly, when we regressed P3b amplitude on our model-based estimates of trial-to-trial efficacy PE, we found a positive relationship (b = 1.23; 95% CrI [0.40, 2.05]; $p_{b<0}$ = 0.01; BF_{10} = 3.10; Figure 3C-left; Table S2).

176

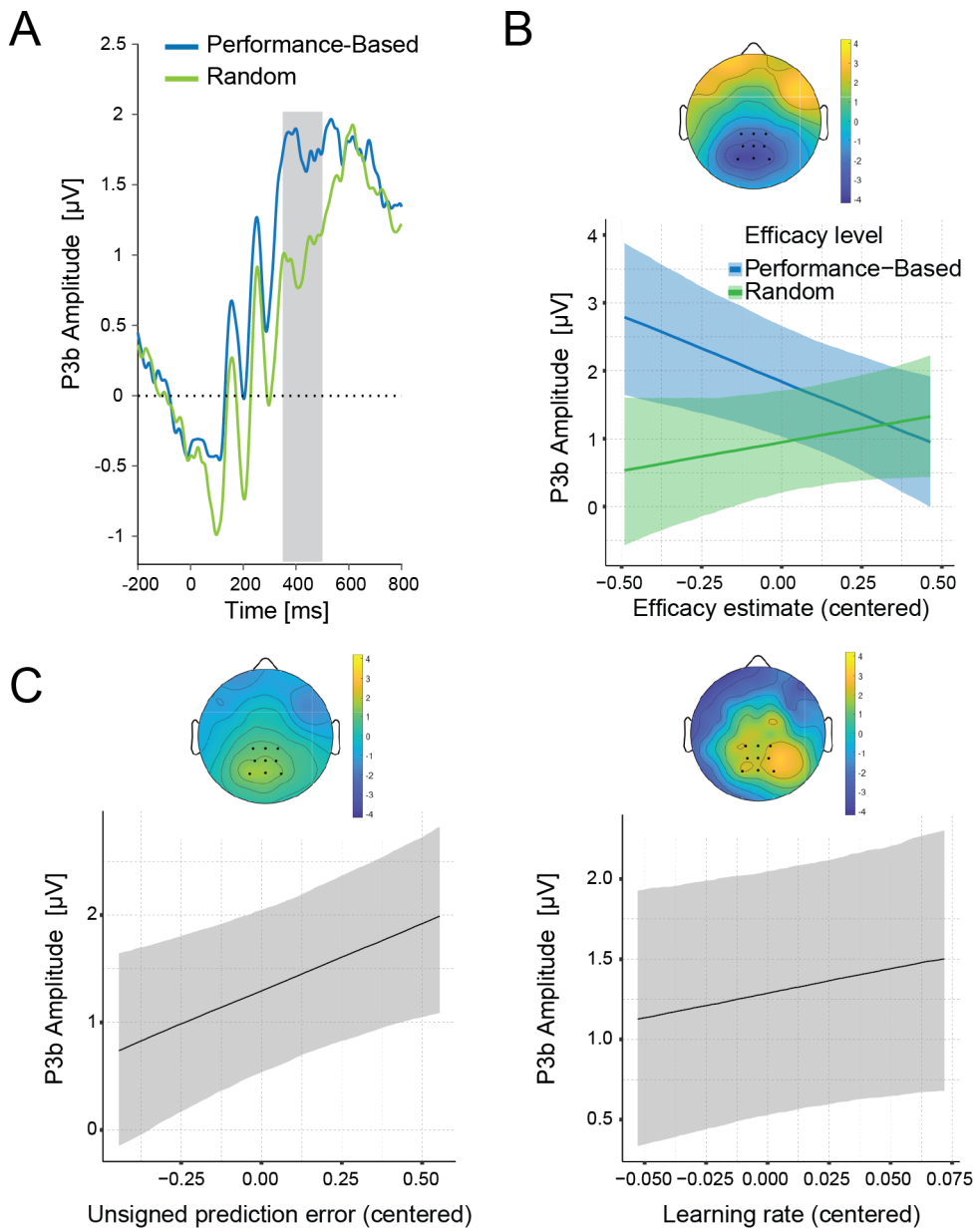

Figure 3. P3b reflects dynamically changing efficacy estimates during processing of efficacy feedback. A. ERP average for the P3b locked to the onset of the efficacy feedback separately for performance-based and random feedback. The grey area shows the time window used for quantifying the P3b. **B.** LMM predicted P3b amplitudes are plotted for performance-based and random feedback as a function of efficacy estimates. The topography shows the interaction of efficacy estimate with efficacy feedback in the P3b time window. **C.** Predicted (centered) effects of unsigned prediction errors (left) and model-based learning rates (right) on the P3b. Shaded error bars represent 95% confidence intervals. Topographies display fixed-effects estimates.

In addition to tracking efficacy PEs, the second key prediction for the P3b if it indexes efficacy learning, is that it should track the extent to which PEs are used to update estimates of efficacy (i.e., the learning rate). In the current study, we found that participants differed in their learning rates for the two forms of efficacy feedback (Performance-Based vs. Random), providing us with a unique opportunity to perform a within-subject test of whether P3b tracked differences in learning rate across these two conditions. Specifically, we could test whether a given subject's P3b was greater in the feedback condition for which they demonstrated a higher learning rate. We found some evidence for this, with the P3b demonstrating an increase for the within-subject feedback condition with the higher learning rate, but this result was not conclusive (b = 3.00; 95% CrI [-0.86, 6.82]; $p_{b<0}$ = 0.10; BF = 1.73; Figure 3C-right). While, theoretically, prediction error and learning rate would interact in predicting the P3b amplitude, we did not observe such an interaction here. The fact that we only found additive effects of prediction errors and learning rate on P3b is, however, in line with previous findings on reward processing (Fischer & Ullsperger, 2013).

The CNV indexes control allocation based on updated expectations of efficacy

Thus far, our findings suggest that participants dynamically updated expectations of their performance efficacy based on feedback, and that the P3b played a role in prediction error-based updating of these expectations. Next, we tested the prediction that learned efficacy estimates determine the expected benefits of control, and thus influence how much control is allocated (Shenhav et al., 2013). We have previously shown that people exert more control when expecting their performance to be more rather than less efficacious on the upcoming trial (Frömer et al., 2021). This was reflected in better behavioral performance and higher amplitudes

of the contingent negative variation (CNV; Figure 3B) - a slow fronto-central waveform preceding target onset, which scales with the amount of control allocated in preparation for the task (Grent-'t-Jong & Woldorff, 2007; Schevernels et al., 2014). Here we used the same marker of control, but, unlike in our previous study, efficacy expectations were (a) learned rather than explicitly instructed; (b) varied over time rather than having a fixed association with a cue; and (c) varied continuously across the range of efficacy probabilities rather than alternating between maximal and minimal efficacy. We were therefore interested in testing whether these dynamically varying expectations of efficacy, as estimated by our model, would still exert the same influence on behavior and neural activity.

Consistent with our predictions and our previous findings, participants tended to perform better when they expected performance to be more efficacious, responding faster on correct trials with increasing model-based estimates of efficacy (Figure 4A; Table S4). This finding replicates the performance effects we observed using instructed cues, albeit not very robustly (b = -11.00; 95% CrI [-23.99, 1.63]; $p_{b>0}$ = 0.08; BF_{01} = 2.55). Like in our previous studies, faster responding was not explained by a change in speed-accuracy trade-off, as accuracy did not decrease with increasing efficacy, but if anything, slightly increased (b = 0.13; 95% CrI [-0.11, 0.38]; $p_{b>0}$ = 0.19; BF_{01} = 2.13). These analyses controlled for the standard behavioral effects related to Stroop congruency (i.e., slower and less accurate responding for incongruent relative to congruent trials; Figure S4).

If the CNV provides an index of control allocation based on current incentive expectations, it should both reflect one's latest efficacy estimate and should predict performance on the upcoming trial. Our findings support both predictions. Regressing single-trial CNV amplitude onto model-based efficacy estimates, and controlling for expectations of reward

(discussed below), we found that these CNV amplitudes tracked current efficacy expectations (b = -0.38; 95% CrI [-0.77, -0.01]; $p_{b>0}$ = 0.05; BF_{01} = 2.93; Figure 4C; Table S5). As with the behavioral finding above, this result provides evidence consistent with our previous CNV finding using instructed cues. We further replicate our earlier finding (Frömer et al., 2021) that larger CNV amplitude in turn predicted better performance on the upcoming trial, with participants responding faster (b = 11.39; 95% CrI [8.64, 14.21]; $p_{b<0}$ = 0; BF_{10} > 100; Figure 4D; Table S6) and more accurately (b = -0.07; 95% CrI [-0.11, -0.02]; $p_{b>0}$ = 0.01; BF_{10} = 3.02; Figure 4E; Table S6) as CNV increased.

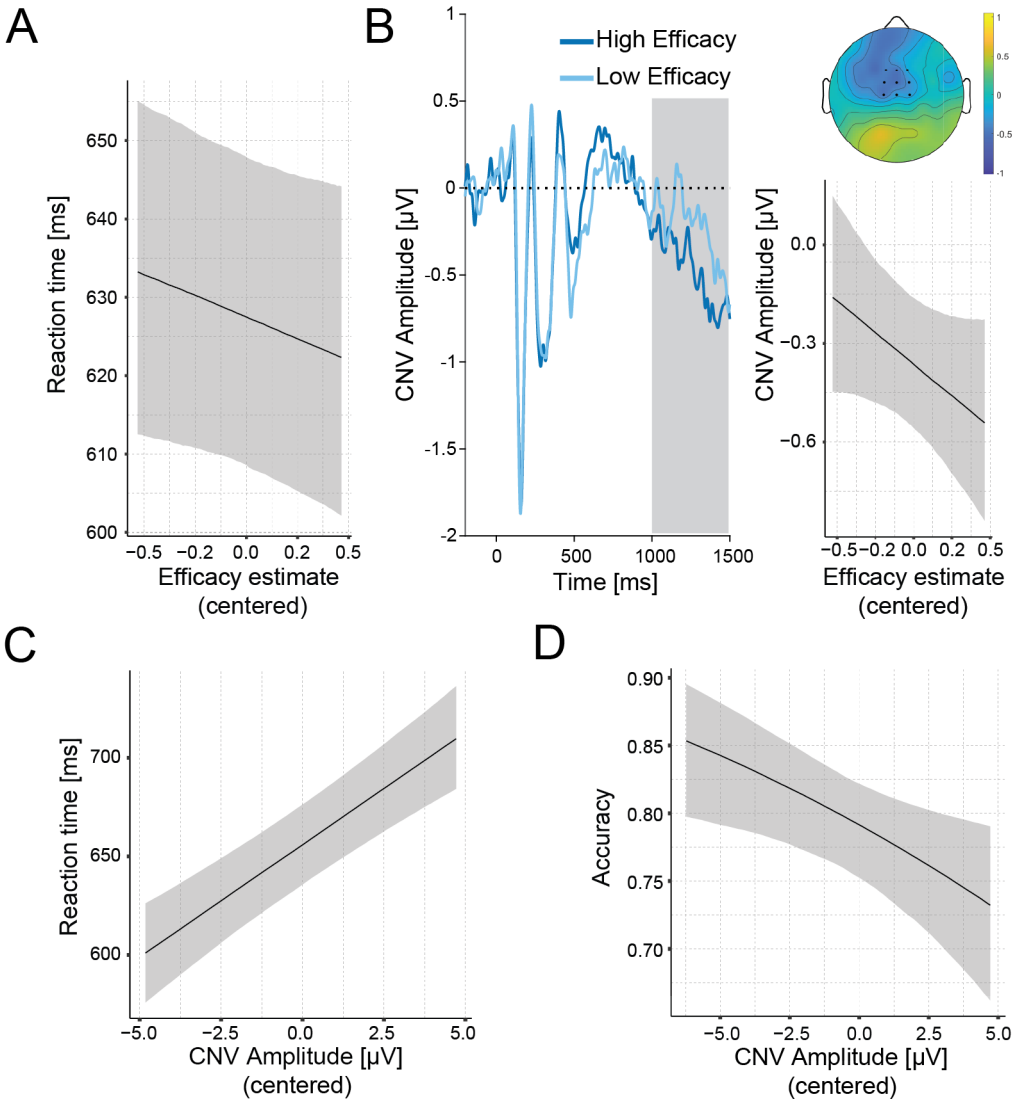

Figure 4. Efficacy estimates influence allocation of control and behavior. A. Higher efficacy predicts faster accurate responses. **B. CNV increases with higher efficacy.** Left: Grand-average ERP for high and low efficacy estimates (median split used for plotting). The shaded area marks the time window used for CNV quantification. Time 0 corresponds to the onset of the neutral cue. Right: Predicted CNV amplitudes as a function of efficacy estimates. The topography shows the fixed effect of the efficacy estimate from the fitted linear model. **C. -D. Larger CNV amplitude predicts better performance.** Predicted accurate RT (**C.**) and accuracy **(D.)** as a function of efficacy estimates. Shaded error bars indicate 95% confidence intervals.

Parallel learning of efficacy and reward rate

We held the amount of reward at stake constant over the course of the experiment, but the frequency of reward (reward rate) varied over the course of the session based on a participant's recent performance, and the participant was sporadically probed for estimates of this reward rate (interleaved with trials that were followed by efficacy probes). Our efficacy manipulation explicitly controlled for this variability by yoking Random-Outcome feedback to a participant's recent reward rate (see methods for details). However, this additional source of variability also provided an opportunity to examine additional mechanisms of learning and adaptation on our task. As in the case of efficacy estimates, reward rate estimates were robustly predicted by reward feedback on previous trials (Table S1; Figure S1A), and this reward rate learning process was well captured by a two learning rate model (cf. Garrett & Daw, 2020; Figure S1B-D). Updates to these reward rate estimates were reflected in P3b modulations of (unsigned) reward prediction errors and associated learning rates (Fischer & Ullsperger, 2013; see Figure S3, Table S3). This pattern of results provides additional evidence that efficacy learning involves similar neural and computational mechanisms as reward-based learning.

Discussion

To evaluate the expected benefits of investing cognitive control into a task, people need to consider whether this control is necessary and/or sufficient for achieving their desired outcome (i.e., whether these efforts are worthwhile). A critical determinant of the worthwhileness of control is performance efficacy, the extent to which performance on a control-demanding task matters for outcome attainment versus those outcomes being determined by performance-unrelated factors. However, the mechanisms through which people estimate the

efficacy of their performance based on previous experience are largely unknown. Here, we identified the computational and neural mechanism through which efficacy is learned and used to guide the allocation of cognitive control. Our results show that participants dynamically updated expectations of the efficacy of their task performance (i.e., the likelihood that this performance will determine reward attainment), and used those expectations to adjust how much control they allocated. The feedback-based updating of efficacy was well-captured by a prediction error-based learning algorithm. Model-based estimates of efficacy and efficacy prediction errors were encoded by canonical neural signatures of effort allocation and belief updating, respectively. This model-based analysis revealed that people allocated more control when they learned that they had more efficacy, extending our previous findings on instructed efficacy (Frömer et al., 2021). Taken together, our results uncover the mechanism through which efficacy estimates are learned and used to inform mental effort investment within a given task environment.

Previous research has characterized the learning algorithms people use to learn the reward value of different states and actions in their environment (Gläscher et al., 2010; Sutton & Barto, 2018). Recent theoretical (Jiang et al., 2014; Lieder et al., 2018; Verbeke & Verguts, 2019) and empirical (Bejjani et al., 2018; Bustamante et al., 2021; Jiang et al., 2020; Otto & Daw, 2019) work has extended this research to show how similar algorithms guide learning and adaptation of cognitive control under varying rewards and task demands within a given task environment. Our findings extend this work further in several ways. First, we show that people leverage weighted prediction errors when learning about the efficacy of task performance, independently of potential reward and task difficulty. Second, we show that they update their efficacy expectations differently depending on whether performance was more efficacious or less efficacious than they expected, demonstrating a striking parallel with dual learning rate models that have been found

to prevail in research on reward learning (Collins & Frank, 2013; Garrett & Daw, 2020; Lefebvre et al., 2017), including in our own reward rate data. Third, we show that participants dynamically adjust their control allocation based on learned efficacy, just as they do for learned rewards and task demands (Bugg et al., 2011; Jiang et al., 2014; Lieder et al., 2018).

Our neural findings build further on past research on learning and control adaptation. The P3b component has been shown to track prediction-error based learning from action-relevant outcome values (Fischer & Ullsperger, 2013; Lohse et al., 2020; Nassar et al., 2019). Here we show that this neural marker tracks learning about efficacy in the same way as it tracks learning about rewards. We found increased P3b amplitudes when people experienced feedback about outcome contingencies that was less expected given their current estimate of efficacy (e.g., expecting low efficacy, but getting performance-contingent feedback), relative to when these contingencies were more expected (e.g., expecting low efficacy and getting random feedback). Our additional finding that P3b amplitude was overall larger for efficacy compared to no efficacy feedback demonstrates that our participants were not just tracking the frequency of the two types of feedback, as predicted by an oddball account. Instead this finding suggests that they were actively learning from the feedback.

Extending previous findings cueing reward and/or efficacy (Frömer et al., 2021; Schevernels et al., 2014), our CNV and behavioral results further show that participants used these learned efficacy estimates to calibrate their effort and their performance. Notably, unlike in previous work, our study shows effort-related changes in CNV amplitude entirely divorced from perceptual cues, providing evidence that this activity truly reflects adjustments in control, rather than reactive processing of features associated with the cue. Taken together, our findings suggest

that similar neural mechanisms underly learning and control adaptation in response to variability in one's efficacy in a task environment, as they do in response to variability in expected rewards.

Our findings build on previous research on how people learn about controllability of their environment. Studies have examined the neural and computational mechanisms by which humans and other animals learn about the contingencies between an action and its associated outcome, and demonstrated that these learned action-outcomes contingencies influence which actions are selected and how vigorously they are enacted (Dickinson & Balleine, 1995; Liljeholm et al., 2011; Ly et al., 2019; Manohar et al., 2017; Moscarello & Hartley, 2017). Outcome contingencies are at least as important to the allocation of cognitive control as they are to action selection, but most control-outcome relationships are significantly more complicated than the types of action-outcome relationships that have previously been studied, where discrete actions typically result in discrete outcomes. Not only are control signals continuous, but they also result in continuous changes in performance, and this performance may or may not determine a particular outcome (Shenhav et al., in press). Learning control-outcome contingencies therefore requires learning about how control states map onto performance (control efficacy) as well as how performance maps onto outcomes (performance efficacy). By describing the mechanisms by which people solve the latter part of this learning problem, and demonstrating that these are comparable to those engaged during action-outcome learning, our study lays critical groundwork for better understanding the links between selection of actions and control states.

Our efficacy-updating results are a reminder that many aspects of feedback are reflected in prediction error signals (Frömer et al., 2020; Langdon et al., 2018). In the present study, we intentionally separated feedback about reward and efficacy to isolate the cognitive and neural learning mechanisms associated with each. In the real world, these aspects are rarely as clearly

separated into different stimuli as they are here. Instead, feedback is multiplexed, and contains information about the values of actions that were taken, as well as about the features and structure of the environment. The same feedback can thus be used to evaluate multiple targets, internal ones, such as the selected response and its predicted outcome, or external ones, such as the source of feedback/environment (Carlebach & Yeung, 2020). Such multi-level prediction error signals might for instance explain, why despite close links between P3b and behavioral adaptation (Chase et al., 2011; Fischer & Ullsperger, 2013; Yeung & Sanfey, 2004), this link is context dependent and attempts to link the P3b uniquely to behavioral updating have failed (Nassar et al., 2019). Reinforcement learning, and predictive inference more generally, have been proposed to not only support the selection of individual actions, but also extended sequences of actions and control signals (Holroyd & Yeung, 2012; Lieder et al., 2018). Alongside evaluations of actions and environmental states, neural signatures of feedback-based learning could thus further reflect evaluations of control signals, their quality or intensity. Given the many potential causes a given outcome can have, and the flexibility that people have in how they use the feedback, it is easy to see how feedback could be misattributed and lead to inaccurate beliefs about performance efficacy. Such beliefs about environmental statistics can drive changes in feedback-processing and behavioral adaptation, above and beyond the true statistics (Schiffer et al., 2017), and are thus of particular importance for understanding some of the cognitive symptoms of mental disorders.

Understanding how efficacy estimates develop based on previous experiences is crucial for understanding why people differ in how much cognitive control they allocate in different situations (Shenhav et al., in press). People differ in their beliefs about how much control they have over their environments (Leotti et al., 2010; Moscarello & Hartley, 2017), and in their

ability to estimate their efficacy (Cohen et al., 2020). Further, many mental disorders, including depression and schizophrenia, have been linked with one's estimates of their ability to control potential outcomes in their environment, including through allocation of control (Huys & Dayan, 2009; Maier & Seligman, 2016), and we have recently proposed that such changes can drive impairments of motivation and control in those populations (Grahek et al., 2019). As we show in this study, when people have learned to expect low efficacy, they will allocate less cognitive control, which can manifest as apparent control deficits. The experimental and modeling approach taken in our study helps uncover a more fine-grained view of how components of motivation are learned and used to support the allocation of cognitive resources. In this way, our study takes a first step toward a better computational and neural account of efficacy learning, which can aid the understanding of individual differences in the willingness to exert mental effort, as well as the development of interventions aimed at teaching individuals when these efforts truly matter.

Methods

Participants

We recruited forty-one participants with normal or corrected-to-normal vision from the Brown University subject pool. One participant was excluded due to technical issues. The final data set included 40 participants (24 females, 16 males; median age = 19). Participants gave informed consent and were compensated with course credits or a fixed payoff of $20. In addition, they received up to $5 as a performance-based bonus ($3.25 on average). The research protocol was approved by Brown University's Institutional Review Board.

Stimuli and task

In the main task, taking approximately 45 minutes, participants performed 288 Stroop trials (Figure 1A). Each trial started with the presentation of a cue (grey circle) that remained on the screen throughout the trial. After a period of 1500 ms, a Stroop stimulus was superimposed until a response was made or 1000 ms elapsed, at which time it was sequentially replaced with two types of feedback presented for 1000 ms each. Each trial onset was preceded by a fixation cross (randomly jittered between 1000 and 1500ms). Participants responded to the ink color of the Stroop stimulus (red, green, blue, or yellow) by pressing one of four keyboard keys (D, F, J, and K). Stimuli were either color words same as the ink color (congruent, n = 108) or different (incongruent, n = 108), or a string of letters "XXXXX" (neutral, n= 72). Feedback informed them whether they obtained a reward (reward, "$50c" or no reward, "$0c") and whether the reward they received depended on their performance (performance-based feedback, a button graphic), or not (random feedback, a dice graphic). The order of the two types of feedback was pseudo-randomized with half of the trials showing reward feedback first and the other half efficacy feedback. Every 2-4 trials the feedback was followed by a probe of efficacy ("How much do you think your rewards currently depend on your performance?") or reward rate ("How often do you think you are currently being rewarded?") to which participants responded on a visual analog scale ranging from 0 to 100. The number and timing of the probes was randomized per subject resulting in a median of 45 efficacy probes (SD=3.38) and 47 reward probes (SD=2.81).

Efficacy (performance-based or random rewards) on each trial was sampled from a binomial distribution with probabilities ranging between 0.1 and 0.9 that drifted over the course of the experiment and were predetermined (Figure 1B). In order to ensure that the performance-

based and random trials did not differ in reward rate, reward feedback for the random trials was sampled from the moving window of the reward feedback of the previous 10 performance-based trials. At the beginning of the experiment a window with 8 rewards and 2 no rewards were used to reflect the calibrated accuracy, and was then updated after every trial to reflect any changes in performance accuracy. Thus, reward rate was not experimentally manipulated in the experiment and remained constant.

Prior to the main task, participants performed several practice phases of the Stroop task (approximately 15 minutes). First, they practiced the mappings between colors and keyboard keys (80 trials). Then they completed a short practice of the Stroop task with written feedback ("correct" or "incorrect") on each trial (30 trials). Participants then completed 100 more of such trials during which we individually calibrated the reaction time deadline such that participants yielded approximately 80% accuracy. In the final practice phase participants were introduced to the two types of feedback which they would see in the main experiment (30 trials).

The experimental task was implemented in Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) for Matlab (MathWorks Inc.) and presented on a 23 inch screen with a 1920 x 1080 resolution. All of the stimuli were presented centrally while the participants were seated 80 cm away from the screen.

Psychophysiological recording and preprocessing

EEG data were recorded at a sampling rate of 500 Hz from 64 Ag/AgCl electrodes mounted in an electrode cap (ECI Inc.), referenced against Cz, using Brain Vision Recorder (Brain Products, München, Germany). Vertical and horizontal ocular activity was recorded from below both eyes (IO1, IO2) and the outer canthi (LO1, LO2), respectively. Impedances were kept below 10 kΩ. Offline, data were processed using custom made Matlab scripts (Frömer et al.,

2018) employing EEGlab functions (Delorme & Makeig, 2004). Data were re-referenced to average reference, ocular artifacts were corrected using brain electric source analyses (BESA; Ille, Berg, & Scherg, 2002) based on separately recorded prototypical eye movements. The cleaned continuous EEG was then high pass filtered at 40 Hz and segmented into epochs around cue onset (-200 to 1500 ms), stimulus onset, and both efficacy and reward feedback (-200 to 800 ms). Baselines were corrected to the average of each 200 ms pre-stimulus interval. Segments containing artifacts, values exceeding ± 150 µV or gradients larger than 50 µV, were excluded from further analyses.

Single trial ERPs were then exported for further analyses in R (R Core Team, 2017). The late CNV was quantified between 1000 and 1500 ms post neutral cue onset (cf. Frömer et al., 2016, 2021; Schevernels et al., 2014) as the average activity over 9 fronto-central electrodes (Fz, F1, F2, FCz, FC1, FC2, Cz, C1, and C2). The P3b was quantified between 350 and 500 ms (cf. Fischer & Ullsperger, 2013) for both reward and efficacy feedback and calculated as the averaged activity over 9 centro-parietal electrodes (Pz, P1, P2, POz, PO1, PO2, CPz, CP1, and CP2).

Learning model and statistical analyses

Learning models. In order to obtain trial-by-trial estimates of efficacy and reward rate, we fitted two temporal difference learning models (Gläscher et al., 2010; Sutton & Barto, 2018) to the continuous (range: 0-1) subjective estimates of efficacy and reward rate (cf. Eldar, Hauser, Dayan, & Dolan, 2016; Nagase et al., 2018; Rutledge et al., 2014). The first model ("1 learning rate model") assumed that the estimate of efficacy for the next trial (Et+1) depended on the current efficacy feedback (Et) and the prediction error (δt) weighted by a constant learning rate (α):

$$E_{t+1} = E_t + \alpha * \delta_t$$

Where $0 \leq \alpha \leq 1$, and the prediction error is calculated as the difference between the contingency feedback on the current trial and the efficacy estimate on that trial: $\delta t = et - Et$. The model started from an initial value (free parameter) and updated the subjective estimate based on the binary feedback on each trial. For example, assuming a learning rate of 0.5 and the initial value of 0.5, the model would update the initial estimate following efficacy feedback signaling "performance-based" (e_t = 1) to 0.75. If on the next trial contingency feedback was "random" (e_t = 0) the subjective estimate would drop to 0.6. Exactly the same model was used to calculate estimates of reward on each trial. The model was fitted separately to the subjective estimates of efficacy and reward rate with only the learning rate as a free parameter. The second model ("2 learning rates model") was the same as the first model, but it included two learning rates: one learning rate for learning from the efficacy feedback, and another for learning from the no efficacy feedback. Finally, as a baseline, we also included the "intercept model" which did not update the efficacy estimate throughout the experiment, but just assumed that the estimate took

one constant value. The models were fit hierarchically to the data using maximum likelihood (for details see the MATLAB scripts available on GitHub).

Statistical analyses. The efficacy and reward rate estimates obtained through fitting the learning model were then used to analyze the behavioral and EEG data. To this end, we fitted Bayesian multilevel regressions to predict subjective estimates of efficacy and reward rates, reaction times, accuracy, as well as the CNV and P3b amplitudes. Subjective estimates of efficacy and reward rate were regressed onto efficacy or reward feedback. Reaction times and accuracies were regressed onto trial-by-trial model-based estimates of efficacy and reward rate, as well as trial-by-trial CNV amplitude, while controlling for congruency. The P3b amplitudes were analyzed in two ways: with trial-by-trial model-based estimates of efficacy and reward rate and current feedback as predictors, and with model-based prediction errors and learning rates for each feedback type. CNV amplitudes were regressed onto trial-by-trial model-based estimates of efficacy. All of the fitted models controlled for the influence of the reward rate estimates. Parallel analyses were done to predict the P3b in response to reward feedback, while controlling for the efficacy estimates.

The regression models were fitted in R with the *brms* package (Bürkner, 2016) which relies on the probabilistic programming language *Stan* (Carpenter et al., 2016) to implement Markov Chain Monte Carlo (MCMC) algorithms and estimate posterior distributions of model parameters. The analyses were done based on the recommendations for Bayesian multilevel modeling using *brms* (Bürkner, 2016, 2017; Nalborczyk & Bürkner, 2019). The fitted models included constant and varying effects (also known as fixed and random) with weakly informative priors (except for the behavioral and CNV analyses, see below for details) for the intercept and the slopes of fixed effects and the likelihoods appropriate for the modeled data (Ex-Gaussian for

reaction times, Bernoulli for accuracy, and Gaussian for the subjective estimates and the EEG amplitudes). The fitted models included all of the fixed effects as varying effects. All of the continuous predictors in the model were centered and the categorical predictors were contrast coded. Four MCMC simulations ("chains"; 20,000 iterations; 19,000 warmup) were run to estimate the parameters of each of the fitted models. The convergence of the models was confirmed by examining trace plots, autocorrelation, and variance between chains (Gelman-Rubin statistic; Gelman & Rubin, 1992). After convergence was confirmed, we analyzed the posterior distributions of the parameters of interest and summarized them by reporting the means of the distribution for the given parameter (*b*) and the 95% credible intervals (95% CrI) of the posterior distributions of that model. We report the proportion of the posterior samples on the relevant side of 0 (e.g., $p_{b<0} = 0.9$), which represents the probability that the estimate is below or above 0. We also report Bayes factors (BF) calculated using the Savage-Dickey method (Wagenmakers et al., 2010). We report the BFs in support of the alternative hypothesis against the null (BF_{10}), except for the analyses of accurate RT and CNV amplitude in which we have informative priors based on our previous study (Frömer, et al., 2020a), and in which case we support the evidence in favor of the null (BF_{10}).

To compare the positive and negative learning rates we fitted a model in which the learning rates were predicted by the learning rate type (Kruschke, 2013). In this model we used Gaussian distributions (mean, standard deviation) as priors (intercept: (0.5,0.5); slopes: (0,0.5)).

We fitted two separate models to predict the subjective estimates of efficacy and reward rate based on previous feedbacks. At each timepoint the estimates were predicted by the current, and previous 4 feedbacks. The feedback on each of the trials (performance-based vs. random or reward vs. no reward) was entered as a constant effect and the models also included the intercept

as a varying effect. As the subjective estimates could vary between 0 and 1, we used Gaussian distributions (intercept: (0.5,0.2); slopes: (0,0.2)) as priors.

For predicting both the P3b amplitude in response to the onset of the efficacy feedback, we fitted two models. First, we fitted a model which included the model-based estimate of efficacy, the actual feedback, and the interaction between the expected efficacy and the observed efficacy feedback. Additionally, we controlled for the reward rate estimate. Second, we fitted separate models which included the model-based prediction errors, the influence of the between-subject learning rates, and their interaction with the prediction errors, while controlling for the estimate of the reward rate. In this analysis the learning rates (one for feedback type for each subject) were mean-centered within subjects and thus any effect of the learning rates is driven by the difference between the random and the performance-based learning rate. For these models we selected wide Gaussian priors (intercept: (5,3); slopes (0,3)). The same logic in building models was applied for the analyses of the reward feedback. In these analyses we focused on the reward feedback processing and how it interacted with the model-based estimates of reward rates, while controlling for the model-based estimates of efficacy. We analyzed only the trials with correct responses for both the efficacy and the reward feedback analyses.

To test the influence of efficacy on the late CNV, we fitted a model which predicted the CNV based on the model-based efficacy estimates, while controlling for the effect of the reward rate estimates. Drawing on the results of our previous study (Frömer, et al., 2020a), this model included Gaussian priors for the intercept (-0.16, 1.59) and the efficacy (-0.30, 0.73) and reward (0,0.73) slopes.

For predicting reaction times and accuracy we fitted models which included congruency (Facilitation: difference between neutral and congruent trials; Interference: difference between

incongruent and neutral trials) and the model-based efficacy estimates, while controlling for the reward rate estimates. We used Gaussian distributions as informative priors based on our previous study (Frömer, et al., 2020a), for both the reaction times (intercept: (624, 58.69); facilitation (15.54, 21.93), interference (61.12, 37.49); efficacy (-10.26, 19.51); reward (0, 19.51)) and accuracy[1] analyses (intercept: (2.11, 0.81); facilitation (-0.45, 0.64), interference (-0.53, 0.81); efficacy (0.09, 0.32); reward (0, 0.32)).

To investigate how the late CNV influences the behavior, we fitted two models in which we predicted the reaction times and accuracy based on the CNV amplitude. The prior distributions for these models were weakly informative Gaussian distributions for predicting both the reaction times (intercept: (650, 200); slope: (0, 50)) and accuracy (intercept: (0.7, 0.2); slope: (0, 0.2)).

To visualize the topographies of the relevant ERP effects, we fitted the relevant models to all 64 channels and then plotted the posterior estimates of the effects of interest at each electrode.

[1] Note that the prior distributions are set in log-odds.

533 **Supplemental Materials**

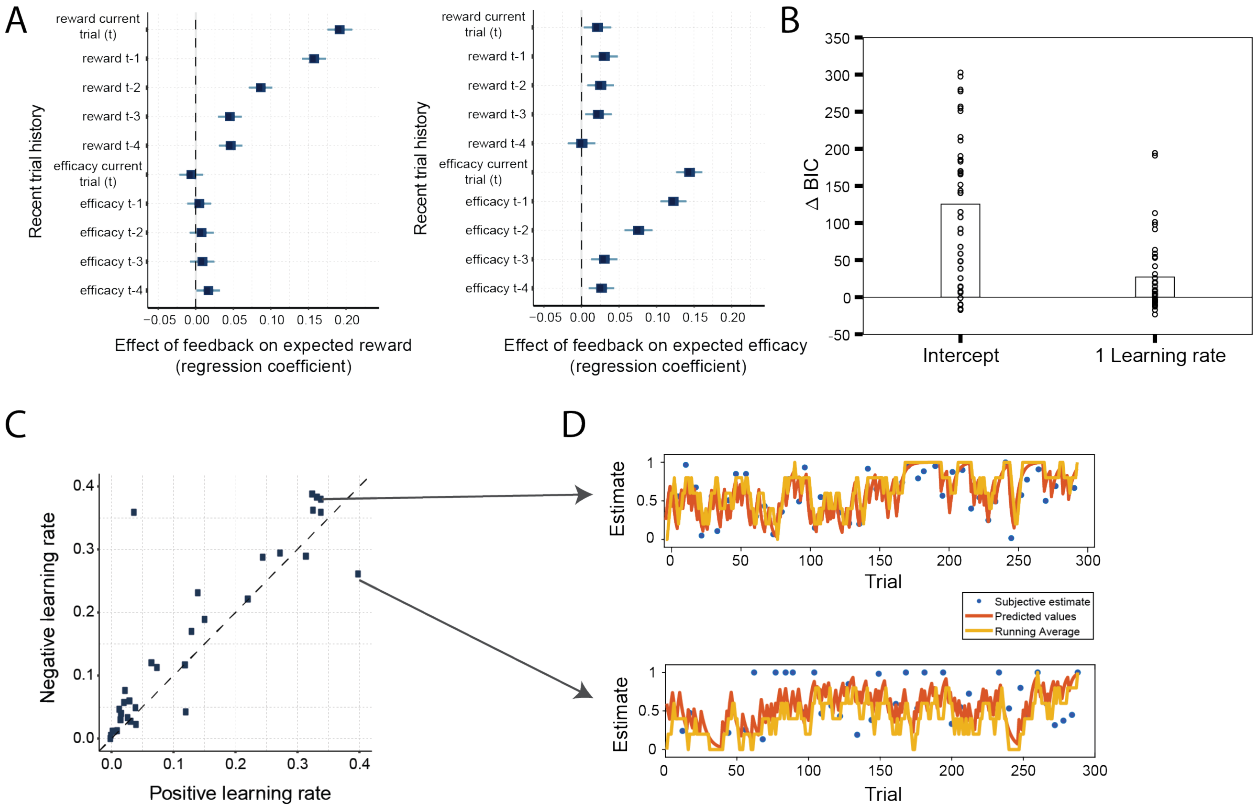

Figure S1. Reward rate learning – model-based analyses. A. Regression weights for the influence of the current and previous feedbacks on the subjective estimates of reward rate (left) and efficacy (right). Reward rate estimates are strongly predicted by the previous reward feedbacks, but not efficacy feedbacks. The reverse is true for the subjective efficacy estimates. Error bars represent 50% and 95% highest density intervals. **B.** Model comparison between the fitted learning models for the reward rate learning model. **C.** Positive and negative learning rate estimates for all subjects for the reward rate learning model. Negative learning rates were numerically, but not statistically larger than positive ones (b = 0.02; 95% CrI [-0.04, 0.08]; $p_{b<0}$ = 0.249; BF_{10} = 0.07) **D.** Subjective and model-based reward rate estimates, and a running average of the previous 5 reward feedbacks, for a subject with a higher learning rate for the no reward feedback (upper) and a subject with the opposite bias (lower).

534

Table S1

Regression weights for the models predicting subjective efficacy and reward rate estimates based on efficacy and reward feedback 5 trials back.

Parameter	*Estimate*	*95% Credible interval*	*Posterior probability (p<0)*	BF_{10}
Subjective efficacy				
Intercept	0.52	0.50, 0.55	0.00	>100
T1 Performance-Based - Random	0.14	0.13, 0.16	0.00	>100
T2 Performance-Based - Random	0.12	0.11, 0.14	0.00	>100
T3 Performance-Based - Random	0.08	0.06, 0.09	0.00	>100
T4 Performance-Based - Random	0.03	0.01, 0.05	0.00	2.18
T5 Performance-Based - Random	0.03	0.01, 0.04	0.00	1.21
T1 Reward - No Reward	0.02	0.00, 0.04	0.03	0.32
T2 Reward - No Reward	0.03	0.01, 0.05	0.00	2.70
T3 Reward - No Reward	0.03	0.01, 0.04	0.01	0.81
T4 Reward - No Reward	0.02	0.00, 0.04	0.02	0.51
T5 Reward - No Reward	0.00	-0.02, 0.02	0.47	0.06
Subjective reward rate				
Intercept	0.50	0.46, 0.53	0.00	>100
T1 Reward - No Reward	0.19	0.18, 0.21	0.00	>100
T2 Reward - No Reward	0.16	0.14, 0.17	0.00	>100
T3 Reward - No Reward	0.09	0.07, 0.10	0.00	>100
T4 Reward - No Reward	0.05	0.03, 0.06	0.00	>100
T5 Reward - No Reward	0.05	0.03, 0.06	0.00	>100
T1 Performance-Based - Random	-0.01	-0.02, 0.01	0.74	0.06
T2 Performance-Based - Random	0.0	-0.01, 0.02	0.32	0.05
T3 Performance-Based - Random	0.01	-0.01, 0.02	0.00	0.05
T4 Performance-Based - Random	0.01	-0.01, 0.02	0.21	0.08
T5 Performance-Based - Random	0.02	0.00, 0.03	0.04	0.22

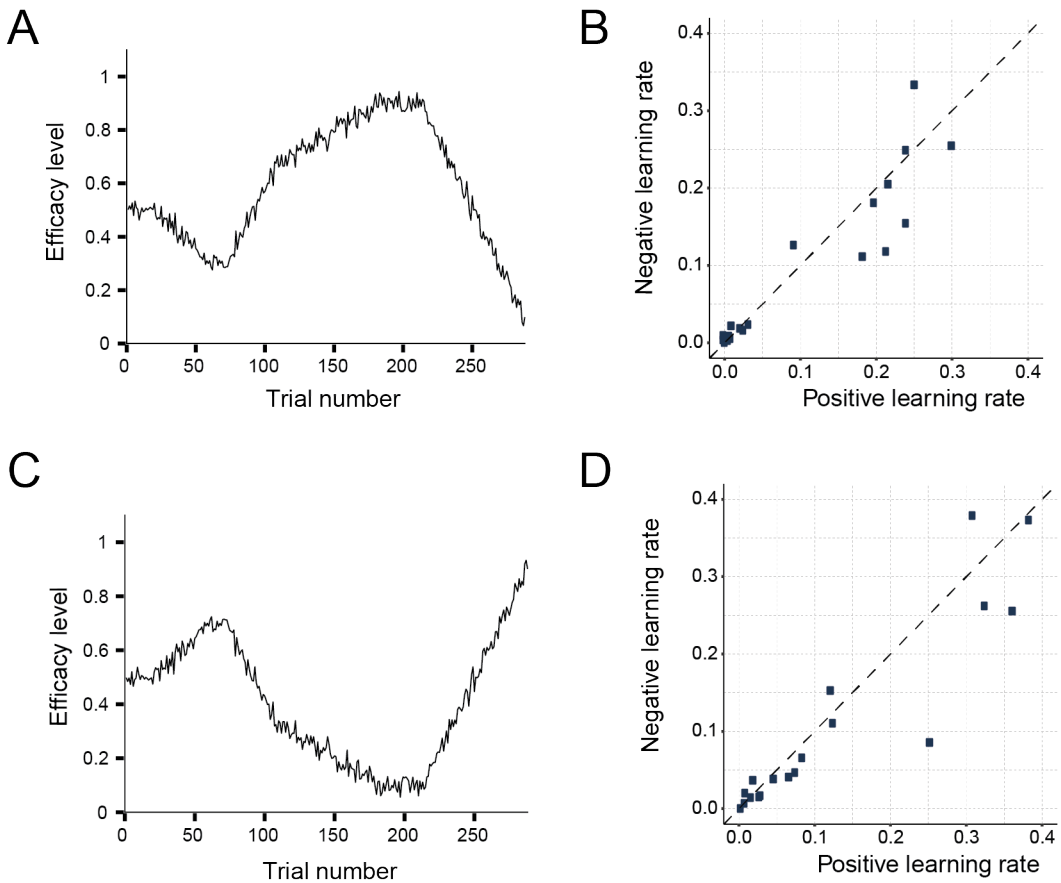

Figure S2. Efficacy learning – the effect of different feedback probability drifts on learning. A. The probability drift of contingent feedback presented to the first half of subjects. **B.** Learning rates for contingent and random feedback for the first half of subjects. Learning rates for the contingent and random feedback did not differ (b = 0.01; 95% CrI [-0.04, 0.08]; $p_{b<0}$ = 0.330; BF_{10} = 0.07). **C.** The inversed probability drift of contingent feedback presented to the second half of participants. **D.** Learning rates for contingent and random feedback did not differ (b = 0.02; 95% CrI [-0.04, 0.08]; $p_{b<0}$ = 0.248; BF_{10} = 0.07).

536

Table S2

Regression weights for the models predicting the P3b to efficacy feedback.

Parameter	*Estimate*	*95% Credible interval*	*Posterior probability (p<0)*	BF_{10}
Efficacy feedback				
Intercept	1.40	0.77, 2.05	0.00	>100
Model-based efficacy	-0.54	-1.46, 0.42	0.84	0.30
Efficacy feedback type (PB-R)	0.90	0.55, 1.24	0.00	>100
Model-based reward rate	-0.27	-1.15, 0.63	0.69	0.20
Model-based efficacy ×Efficacy feedback type	-2.77	-4.08, -1.48	1.00	>100
Model-based efficacy slope for Performance-Based feedback	-1.93	-3.07, -0.75	1.00	6.04
Model-based efficacy slope for Random feedback	0.84	-0.28, 2.00	0.11	0.43
Predictions errors and learning rates				
Intercept	1.30	0.69, 1.92	0.00	>100
Unsigned prediction error	1.23	0.40, 2.05	0.01	3.10
Learning rate	3.00	-0.86, 6.82	0.10	1.73
Model-based reward rate	-0.22	-1.16, 0.71	0.66	0.21
Unsigned prediction error ×Learning rate	0.51	-4.37, 5.30	0.43	0.96

537

ERP analyses of the reward rate feedback

To investigate the effect of the model-based reward rate estimates on the processing of reward, we performed complementary analyses on the P3b related to the processing of the reward feedback (Figure S3A). We expected that a similar learning mechanism should operate on both reward and efficacy feedback, but that the neural markers of feedback processing should be sensitive to different model-based estimates. Thus, we expected the reward-locked P3b to be sensitive to reward-rate estimates, but not to efficacy estimates.

Our results showed larger P3b amplitudes to no reward compared to reward feedback (note that reward rate was approximately .80 and negative feedback less expected overall), and the processing of the reward feedback was influenced by reward-rate estimates, although to less extent than for the P3b (Figure S3B; Table S3). Probing the effects of learning directly, we found robust effects of the prediction errors and learning rates (Figure S3C). The P3b to reward feedback was not influenced by the model-based efficacy estimates in either of these analyses.

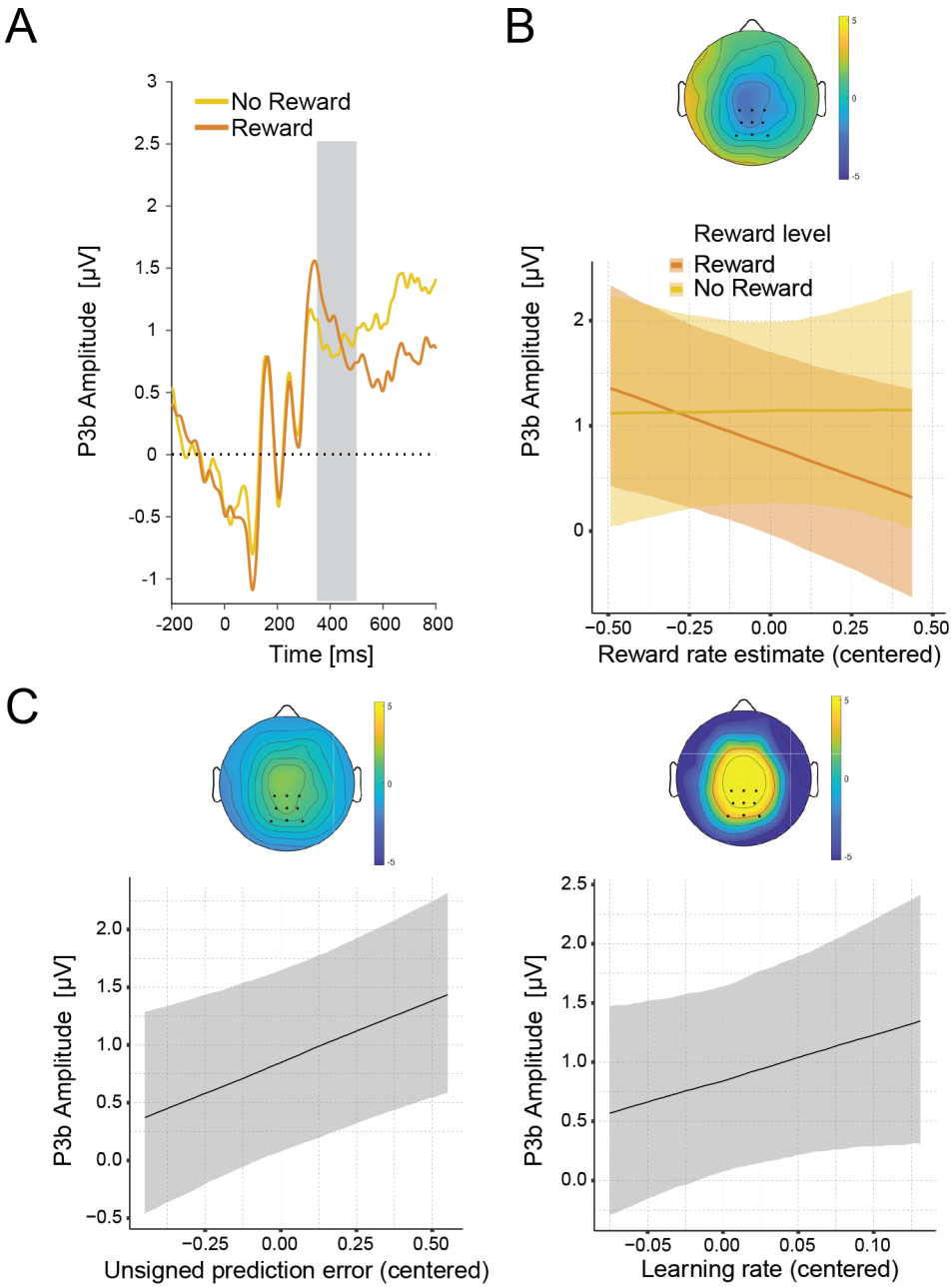

Figure S3. Learning reward rate estimate from reward feedback. A. ERP average for the P3b locked to the onset of reward feedback. **B.** Interaction between reward feedback and the model-based reward rate estimate. **C.** The effects of unsigned prediction errors (left) and learning rates (right) on the P3b.

552

Table S3

Regression weights for the models predicting the P3b to reward feedback.

Parameter	*Estimate*	*95% Credible interval*	*Posterior probability (p<0)*	*BF_{10}*
Efficacy feedback				
Intercept	0.98	0.30, 1.66	0.01	>100
Model-based efficacy	-0.54	-1.33, 0.25	0.87	0.30
Reward feedback type (R-NR)	-0.32	-0.76, 0.12	0.90	0.19
Model-based reward rate	0.05	-0.55, 0.66	0.44	0.12
Model-based reward ×Reward feedback type	-1.14	-2.57, 0.30	0.90	0.65
Model-based efficacy slope for No Reward feedback	0.03	-1.25, 1.28	0.99	1.54
Model-based efficacy slope for Reward feedback	-1.11	-1.91, -0.27	0.48	0.23
Predictions errors and learning rates				
Intercept	0.87	0.22, 1.52	0.02	>100
Unsigned prediction error	1.06	0.47, 1.64	0.00	7.55
Learning rate	3.78	-0.76, 8.12	0.08	2.79
Model-based efficacy	-0.03	-0.61, 0.57	0.54	0.12
Unsigned prediction error ×Learning rate	0.37	-4.44, 5.20	0.45	0.99

553

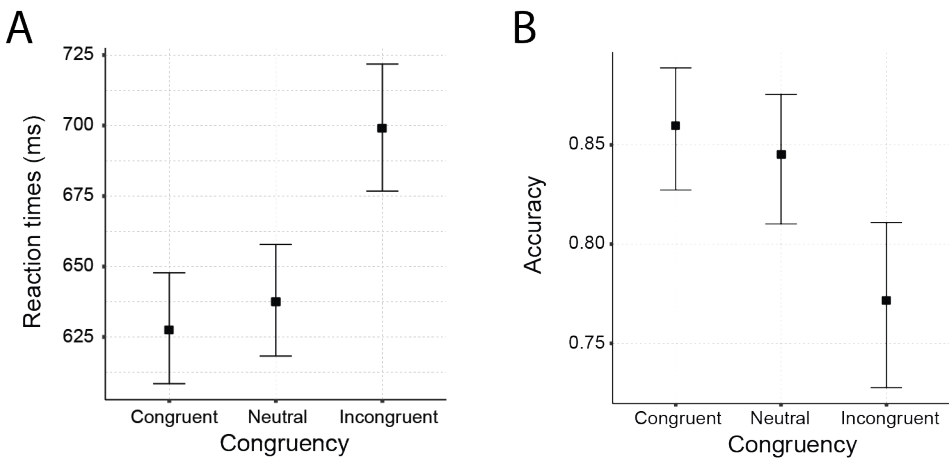

Figure S4. Behavioral effects of congruency. A. In comparison to the neutral trials, participants were faster to respond to congruent (b = 9.93; 95% CrI [4.26, 15.51]; $p_{b<0}$ = 0.002; BF = 499) and slower to respond to incongruent trials (b = 60.90; 95% CrI [51.43, 70.49]; $p_{b<0}$ = 0; BF > 4000). **B.** In comparison to the neutral trials, participants were equally likely to be correct when responding to congruent (b = -0.11; 95% CrI [-0.25, 0.03]; $p_{b>0}$ = 0.096; BF = 9.42) and less likely to be correct when responding to incongruent trials (b = -0.48; 95% CrI [-0.61, -0.34]; $p_{b>0}$ = 0; BF > 4000). The regression

Table S4

Regression weights for the models predicting reaction times and accuracy based on efficacy and reward rate estimates.

Parameter	*Estimate*	*95% Credible interval*	*Posterior probability (p<0)*	BF_{01}
Reaction times				
Intercept	654.88	638.52, 671.45	0.00	>100
Facilitation	9.95	4.56, 15.42	0.00	1.64
Interference	61.44	51.94, 70.85	0.00	6.54
Model-based efficacy	-11.00	-23.99, 1.63	0.92	2.55
Model-based reward rate	-9.99	-26.81, 9.08	0.83	0.98
Accuracy				
Intercept	1.58	1.39, 1.77	0.00	>100
Facilitation	-0.12	-0.26, 0.03	0.91	0.01
Interference	-0.48	-0.61, -0.35	1.00	8.41
Model-based efficacy	0.13	-0.11, 0.38	0.19	2.13
Model-based reward rate	-0.19	-0.45, 0.07	0.89	0.89

Table S5

Regression weights for the models predicting the CNV based on efficacy and reward rate.

Parameter	*Estimate*	*95% Credible interval*	*Posterior probability (p<0)*	BF_{01}
CNV				
Intercept	-0.36	-0.52, -0.19	1.00	>100
Model-based efficacy	-0.38	-0.77, -0.01	0.95	2.93
Model-based reward rate	-0.06	-0.47, 0.35	0.60	2.90

Table S6

Regression weights for the models predicting reaction times and accuracy based on the CNV.

Parameter	*Estimate*	*95% Credible interval*	*Posterior probability (p<0)*	BF_{10}
Reaction times				
Intercept	656.03	639.14, 673.19	0.00	>100
CNV amplitude	11.39	8.64, 14.21	0.00	>100
Accuracy				
Intercept	1.33	1.15, 1.50	0.00	>100
CNV amplitude	-0.07	-0.11, -0.02	0.99	3.02

Acknowledgements

This work was supported by the Special Research Fund (BOF) of Ghent University [grant #01D02415] (I.G.) the Research Foundation Flanders (FWO) travel grant [grant #V432718N] (I.G.), a Center of Biomedical Research Excellence grant P20GM103645 from the National Institute of General Medical Sciences (A.S.), and the Alfred P. Sloan Foundation Research Fellowship in Neuroscience (A.S.). The funding sources were not involved in the study design; collection, analysis, and interpretation of data; writing of the report; and decision to submit the article for publication. We would like to thank Natalie Knowles and Hattie Xu for their help with data collection, Carolyn Dean Wolf for her help with programming the task, and Harrison Ritz for his help with constructing the efficacy drifts.

Data availability

Raw and pre-processed data, materials, and analysis scripts will be available on Github and OSF upon publication.

References

Bejjani, C., Zhang, Z., & Egner, T. (2018). Control by association: Transfer of implicitly primed attentional states across linked stimuli. *Psychonomic Bulletin and Review*, *25*(2), 617–626. https://doi.org/10.3758/s13423-018-1445-6

Botvinick, M. M., & Cohen, J. D. (2014). The computational and neural basis of cognitive control: Charted territory and new frontiers. *Cognitive Science*, *38*(6), 1249–1285. https://doi.org/10.1111/cogs.12126

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

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. https://doi.org/10.1037/a0019957

Bürkner, P.-C. (2016). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Bürkner, P.-C. (2017). Advanced Bayesian Multilevel Modeling with the R Package brms. *ArXiv:1705.11123*.

Bustamante, L., Lieder, F., Musslick, S., Shenhav, A., & Cohen, J. (2021). Learning to Overexert Cognitive Control in a Stroop Task. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-020-00845-x

Carlebach, N., & Yeung, N. (2020). Flexible use of confidence to guide advice requests. *PsyArXiv*.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, *2*(20), 1–37.

Chambon, V., Théro, H., Vidal, M., Vandendriessche, H., Haggard, P., & Palminteri, S. (2020). Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-020-0919-5

Chase, H. W., Swainson, R., Durham, L., Benham, L., & Cools, R. (2011). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *Journal of Cognitive Neuroscience*, *23*(4), 936–946. https://doi.org/10.1162/jocn.2010.21456

Chiu, Y. C., & Egner, T. (2019). Cortical and subcortical contributions to context-control learning. *Neuroscience and Biobehavioral Reviews*, *99*(December 2018), 33–41. https://doi.org/10.1016/j.neubiorev.2019.01.019

Cohen, A., Nussenbaum, K., Dorfman, H. M., Gershman, S. J., & Hartley, C. A. (2020). The rational use of causal inference to guide reinforcement learning strengthens with age. *Npj Science of Learning*.

Collins, Anne G E, & Frank, M. J. (2014). Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological Review*, *121*(3), 337–366. https://doi.org/10.1037/a0037015

Collins, Anne Gabrielle Eva, & Frank, M. J. (2013). Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychological Review*, *120*(1), 190–229. https://doi.org/10.1037/a0030852

Daw, N. D., & Tobler, P. N. (2013). Value Learning through Reinforcement: The Basics of Dopamine and Reinforcement Learning. In *Neuroeconomics: Decision Making and the Brain: Second Edition* (pp. 283–298). https://doi.org/10.1016/B978-0-12-416008-8.00015-2

Delorme, A., & Makeig, S. (2004). EEGLAB: an open sorce toolbox for analysis of single-trail EEG dynamics including independent component anlaysis. *Journal of Neuroscience Methods*, *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dickinson, A., & Balleine, B. (1995). Motivational Control of Instrumental Action. *Current Directions in Psychological Science*, *4*(5), 162–167. https://doi.org/10.1111/1467-8721.ep11512272

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312–325. https://doi.org/10.1016/j.neuron.2013.09.007

Eldar, E., Hauser, T. U., Dayan, P., & Dolan, R. J. (2016). Striatal structure and function predict individual biases in learning to avoid pain. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(17), 4812–4817. https://doi.org/10.1073/pnas.1519829113

Fischer, A. G., & Ullsperger, M. (2013). Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron*, *79*(6), 1243–1255. https://doi.org/10.1016/j.neuron.2013.07.006

Frömer, R., Lin, H., Dean Wolf, C. K., Inzlicht, M., & Shenhav, A. (2021). Expectations of reward and efficacy guide cognitive control allocation. *Nature Communications*, *12*(1), 1030. https://doi.org/10.1038/s41467-021-21315-z

Frömer, R., Maier, M., & Rahman, R. A. (2018). Group-level EEG-processing pipeline for flexible single trial-based analyses including linear mixed models. *Frontiers in Neuroscience*, *12*(48). https://doi.org/10.3389/fnins.2018.00048

Frömer, R., Nassar, M., Stürmer, B., Sommer, W., & Yeung, N. (2020). I knew that! Response-based OutcomePredictions and Confidence Regulate Feedback Processing and Learning.

BioRxiv.

Frömer, R., Stürmer, B., & Sommer, W. (2016). (Don't) Mind the effort: Effects of contextual interference on ERP indicators of motor preparation. *Psychophysiology*, *53*(10), 1577–1586. https://doi.org/10.1111/psyp.12703

Garrett, N., & Daw, N. D. (2020). Biased belief updating and suboptimal choice in foraging decisions. *Nature Communications*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-16964-5

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–595. https://doi.org/10.1016/j.neuron.2010.04.016

Grahek, I., Shenhav, A., Musslick, S., Krebs, R. M., & Koster, E. H. W. (2019). Motivation and cognitive control in depression. *Neuroscience & Biobehavioral Reviews*, *102*(March), 371–381. https://doi.org/10.1016/j.neubiorev.2019.04.011

Grent-'t-Jong, T., & Woldorff, M. G. (2007). Timing and sequence of brain activity in top-down control of visual-spatial attention. *PLoS Biology*, *5*(1), 0114–0126. https://doi.org/10.1371/journal.pbio.0050012

Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, *16*(2), 122–128. https://doi.org/10.1016/j.tics.2011.12.008

Huys, Q. J. M., & Dayan, P. (2009). A Bayesian formulation of behavioral control. *Cognition*, *113*(3), 314–328. https://doi.org/10.1016/j.cognition.2009.01.008

Ille, N., Berg, P., & Scherg, M. (2002). Artifact Correction of the Ongoing EEG Using Spatial Filters Based on Artifact and Brain Signal Topographies. *Journal of Clinical Neurophysiology*, *19*(2), 113–124. https://doi.org/10.1097/00004691-200203000-00002

Jiang, J., Bramão, 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. *Journal of Neuroscience*, *40*(10), 2343–2356. https://doi.org/10.1523/JNEUROSCI.1869-19.2020

Jiang, J., Heller, K., & Egner, T. (2014). Bayesian modeling of flexible cognitive control. *Neuroscience and Biobehavioral Reviews*, *46*(P1), 30–43. https://doi.org/10.1016/j.neubiorev.2014.06.001

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3. *Perception 36 ECVP Abstract Supplement*.

Kool, W., & Botvinick, M. (2018). Mental labour. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-018-0401-9

Kruschke, J. K. (2013). Bayesian estimation supersedes the T test. *Journal of Experimental Psychology: General*, *142*(2), 573–588. https://doi.org/10.1037/a0029177

Langdon, A. J., Sharpe, M. J., Schoenbaum, G., & Niv, Y. (2018). Model-based predictions for dopamine. *Current Opinion in Neurobiology*, *49*, 1–7. https://doi.org/10.1016/j.conb.2017.10.006

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., & Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, *1*(4), 1–9. https://doi.org/10.1038/s41562-017-0067

Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose: The origins and value of

the need for control. *Trends in Cognitive Sciences*, *14*(10), 457–463. https://doi.org/10.1016/j.tics.2010.08.001

Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. *Plos Computational Biology*, *14*(4). https://doi.org/10.13140/RG.2.2.24500.14721

Liljeholm, M., Tricomi, E., O'Doherty, J. P., & Balleine, B. W. (2011). Neural Correlates of Instrumental Contingency Learning: Differential Effects of Action-Reward Conjunction and Disjunction. *Journal of Neuroscience*, *31*(7), 2474–2480. https://doi.org/10.1523/JNEUROSCI.3354-10.2011

Lohse, K. R., Miller, M. W., Daou, M., Valerius, W., & Jones, M. (2020). Dissociating the contributions of reward-prediction errors to trial-level adaptation and long-term learning. *Biological Psychology*, *149*(January 2019). https://doi.org/10.1016/j.biopsycho.2019.107775

Ly, V., Wang, K. S., Bhanji, J., & Delgado, M. R. (2019). A Reward-Based Framework of Perceived Control. *Frontiers in Neuroscience*, *13*(February), 1–11. https://doi.org/10.3389/fnins.2019.00065

Maier, S. F., & Seligman, M. E. P. (2016). Learned helplessness at fifty: Insights from neuroscience. *Psychological Review*, *123*(4), 1–19. https://doi.org/10.1037/rev0000033

Manohar, S. G., Chong, T. T. J., Apps, M. A. J., Batla, A., Stamelou, M., Jarman, P. R., Bhatia, K. P., & Husain, M. (2015). Reward Pays the Cost of Noise Reduction in Motor and Cognitive Control. *Current Biology*, *25*(13), 1707–1716. https://doi.org/10.1016/j.cub.2015.05.038

Manohar, S. G., Finzi, R. D., Drew, D., & Husain, M. (2017). Distinct Motivational Effects of

Contingent and Noncontingent Rewards. *Psychological Science*, *28*(7), 1016–1026. https://doi.org/10.1177/0956797617693326

Moscarello, J. M., & Hartley, C. A. (2017). Agency and the Calibration of Motivated Behavior. *Trends in Cognitive Sciences*, *21*(10), 725–735. https://doi.org/10.1016/j.tics.2017.06.008

Nagase, A. M., Onoda, K., Foo, J. C., Haji, T., Akaishi, R., Yamaguchi, S., Sakai, K., & Morita, K. (2018). Neural mechanisms for adaptive learned avoidance of mental effort. *Journal of Neuroscience*, *38*(10), 2631–2651. https://doi.org/10.1523/JNEUROSCI.1995-17.2018

Nalborczyk, L., & Bürkner, P.-C. (2019). An Introduction to Bayesian Multilevel Models Using brms: A Case Study of Gender Effects on Vowel Variability in Standard Indonesian. *Journal of Speech, Language, and Hearing Research*.

Nassar, M. R., Bruckner, R., & Frank, M. J. (2019). Statistical context dictates the relationship between feedback-related EEG signals and learning. *ELife*, *8*, 1–26. https://doi.org/10.7554/eLife.46975

Nigbur, R., Schneider, J., Sommer, W., Dimigen, O., & Stürmer, B. (2015). Ad-hoc and context-dependent adjustments of selective attention in conflict control: An ERP study with visual probes. *NeuroImage*, *107*, 76–84. https://doi.org/10.1016/j.neuroimage.2014.11.052

Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *Journal of Neuroscience*, *32*(2), 551–562. https://doi.org/10.1523/JNEUROSCI.5498-10.2012

Niv, Y., Joel, D., & Dayan, P. (2006). A normative perspective on motivation. *Trends in Cognitive Sciences*, *10*(8), 375–381. https://doi.org/10.1016/j.tics.2006.06.010

Otto, A. R., & Daw, N. D. (2019). The opportunity cost of time modulates cognitive effort. *Neuropsychologia*, *123*(October 2017), 92–105.

https://doi.org/10.1016/j.neuropsychologia.2018.05.006

Pelli, D. D. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

R Core Team. (2017). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing.

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, *111*(33), 12252–12257. https://doi.org/10.1073/pnas.1407535111

Schevernels, H., Krebs, R. M., Santens, P., Woldorff, M. G., & Boehler, C. N. (2014). Task preparation processes related to reward prediction precede those related to task-difficulty expectation. *NeuroImage*, *84*, 639–647. https://doi.org/10.1016/j.neuroimage.2013.09.039

Schiffer, A. M., Siletti, K., Waszak, F., & Yeung, N. (2017). Adaptive behaviour and feedback processing integrate experience and instruction in reinforcement learning. *NeuroImage*, *146*, 626–641. https://doi.org/10.1016/j.neuroimage.2016.08.057

Shenhav, A., Botvinick, M., & Cohen, J. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

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*(1), 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Shenhav, A., Prater Fahey, M., & Grahek, I. (n.d.). *Decomposing the motivation to exert mental effort*.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press.

Verbeke, P., & Verguts, T. (2019). Learning to synchronize: How biological agents can couple neural task modules for dealing with the stability-plasticity dilemma. *PLoS Computational Biology*, *15*(8), 1–25. https://doi.org/10.1371/journal.pcbi.1006604

Verguts, T., Vassena, E., & Silvetti, M. (2015). Adaptive effort investment in cognitive and physical tasks: a neurocomputational model. *Frontiers in Behavioral Neuroscience*, *9*(March). https://doi.org/10.3389/fnbeh.2015.00057

Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *Journal of Neuroscience*, *24*(28), 6258–6264. https://doi.org/10.1523/JNEUROSCI.4537-03.2004