

# Reward reinforcement creates habitual selection of goals

Ian C. Ballard<sup>1</sup>, Michael Waskom<sup>2</sup>, Kerry C. Nix<sup>3</sup>, & Mark D'Esposito<sup>1</sup>

<sup>1</sup>Psychology Department, University of California, Riverside

<sup>2</sup>Center for Neural Science, New York University

<sup>3</sup>Department of Neuroscience, University of Pennsylvania

*Correspondence:* ianballard@gmail.com

## ABSTRACT

Stimulus-response habits benefit behavior by automatizing the selection of rewarding actions. However, this automaticity comes at the cost of reduced flexibility to adapt behavior when circumstances change. The goal-directed system is thought to counteract the habit system by providing the flexibility to pursue context-appropriate behaviors. The dichotomy between habitual action selection and flexible goal-directed behavior has recently been challenged by findings showing that rewards bias both action and goal selection. Here, we test whether reward reinforcement can give rise to habitual goal selection much as it gives rise to habitual action selection. We designed a rewarded, context-based perceptual discrimination task in which performance on one rule was reinforced. Using drift-diffusion models and psychometric analyses, we found that reward facilitates the selection and execution of rules while also impairing the ability to switch to alternative, low-reward rules. Strikingly, we found that these biases persisted in a test phase in which rewards were no longer available. This facilitation of the performance of high-reward rules, at the cost of reduced cognitive flexibility, demonstrates that reward can give rise to habitual rule selection. Our findings build on recent work showing a role for the reward learning system in influencing how the goal-directed system selects goals. Moreover, they demonstrate that reward creates lasting biases in goal-directed behavior that could contribute to cognitive inflexibility in healthy and clinical populations.

## INTRODUCTION

Habits are powerful determinants of daily decisions and contribute to maladaptive behaviors in neurocognitive disorders (Lhermitte, 1983; Wood & Runger, 2016). Habitual behavior is often characterized as a rote or automatic behavioral response to a specific stimulus, such as stopping at a red light (Knowlton et al., 1996; Schneider & Shiffrin, 1977). However, many habits operate at the level of goals rather than specific actions. For example, someone who has a habit of exercising will habitually pursue exercise-related behaviors, such as navigating to a gym or researching exercise-relevant information. In both of these cases, the pursuit of a goal (stopping at a red light or exercising) is beneficial; however, in the former case, a specific action, pressing the brake pedal, achieves the goal, whereas in the latter case, a variety of context-dependent strategies are useful for goal pursuit. The concept of a “goal habit” postulates that the selection of a goal state is influenced by reward learning (Cushman & Morris, 2015), and flexible cognitive control strategies are deployed to pursue these goals. Maladaptive compulsions in clinical contexts often involve habitual activation of goals. For example, a person suffering from drug addiction may exhibit goal habits, such as exploring novel strategies for attaining drugs, and stimulus-response habits, such as drug-cue-induced approach behavior (Vandaele & Ahmed, 2021). Recent research has emphasized the role of the habit system in driving stereotyped mental behaviors in anxiety (Brewer & Roy, 2021), anorexia nervosa (Steinglass & Walsh, 2006), obsessive-compulsive disorders (Gillan & Robbins, 2014; Voon et al., 2015), and in Parkinson’s disease (Weintraub, 2008). However, the neural and psychological mechanisms underlying goal habits remain underspecified.

Habitual action selection is thought to arise in part from the dopaminergic adjustment of corticostriatal synaptic strength (DeLong, 1990; Graybiel, 1998; Niv, 2009). In response to reward, dopamine release strengthens the corticostriatal synapses of cortical pools representing a chosen action. This corticostriatal plasticity favors the future selection of actions that lead to rewards. Neurons in the lateral prefrontal cortex represent abstract rules and goals (Wallis et al., 2001)) rather than actions but share a similar, overlapping corticostriatal architecture with motor cortex (Alexander et al., 1986; Haber, 2012). It has been hypothesized that reward reinforces abstract task representations analogously to cortical action representations (Badre & Frank, 2012; Collins & Frank, 2013; Frank & Badre, 2012; Radulescu et al., 2019; Ribas-Fernandes et al., 2011). Recent research has confirmed key predictions of this model by showing that reward history influences the selection of goal states (Cushman & Morris, 2015) and hierarchically

structured task sets (Collins & Frank, 2013; Eckstein & Collins, 2020; Rmus et al., 2021). The present study builds upon this work by testing whether reward reinforcement of abstract representations causes goal habits.

We designed a behavioral experiment to test three key three predictions of the goal habit model. First, execution of habitual goals ought to be improved relative to other goals. Second, the ability to adapt goals under changing contexts should be reduced. Third, habitual goal selection should persist even after the conditions which gave rise to the habit have changed. A key feature of goals is that they guide context dependent behavior, and we operationalized this context-dependence using a rule-based perceptual discrimination task. We found that reward reinforcement of rules influenced behavior in a manner consistent with all three predictions: Execution of the high-reward rule was improved, the ability to adapt behavior away from the high-reward rule was reduced, and both of these effects persisted after the opportunity to earn rewards was eliminated. These results show that reward induces habitual biases in goal-directed behavior.

## METHODS

### ***Subject details***

The study design and methods were approved by and followed the ethical procedures of the University of California, Berkeley Committee for the Protection of Human Subjects. Eighty-six subjects provided written informed consent, 65 females, median age 20 years, SD 4.83, range 18 - 51. Data from the test blocks are missing from one subject due to a computer error. The target sample size, eighty-five subjects, was chosen to have 80% power to detect a medium-sized correlation ( $r = 0.3$ ) at an alpha of 0.05. Because we did not identify any outlier subjects in behavioral performance (defined as three standard deviations below the mean accuracy), and all subjects performed well-above chance, no subjects were excluded.

### ***Task design***

Subjects performed a context-based perceptual discrimination task in which they could earn rewards for accurate performance (M. L. Waskom et al., 2019; M. L. Waskom & Wagner, 2017). On each trial, subjects responded based on one of three rules, color, shape, or motion

direction of a field of colored, moving shapes. The dots could be primarily pink or green, primarily circles or crosses, and moving primarily up or down. Dominant color, shape, and motion direction were balanced across each run. Subjects were given up to two seconds to respond using the '1' and '2' keys on a standard keyboard and could respond at any time during the stimulus period. The stimulus remained on the screen for two seconds regardless of when the subject responded. All three rules shared the same keys, i.e., response '1' could signal 'green' on a color trial and 'up' on a motion trial. The rule indicating which dimension to respond to was cued simultaneously with stimulus onset by a three-to-five-sided polygon drawn at the center of the stimulus array. The assignment between shape cue and rule remained consistent throughout the study for each participant and was counterbalanced across participants.

Coherence varied pseudorandomly across trials and independently across the three dimensions of each stimulus. Coherence varied in four evenly-spaced steps from hardest (least coherent) to easiest (most coherent), with color and shape coherence ranging from 0.52 to 0.64 (zero coherent information is 0.50) and motion coherence ranging from 0.02 to 0.14 (zero coherent information is 0.0). These levels were chosen based on piloting to provide a range in performance from slightly-above chance accuracy to near-ceiling accuracy on all three rules. There were differences in accuracy between the rules,  $F(1.73, 147) = 20$ ,  $\eta^2g = .059$ ,  $p < .001$ , shape  $M$ : 72.2%, motion  $M$ : 74.1%, color  $M$ : 77.1%; however, there were no differences in RT,  $p > .2$ . Although this accuracy difference contributes noise to our data, rule counterbalancing was designed to prevent any systematic influence on the reported results.

The task was organized into a reward phase and a test phase. During the reward phase, subjects were instructed that some trials carried the potential to earn rewards for correct responses. Incorrect responses prevented the subject from earning a reward. One rule was randomly chosen for each subject to be the high-reward rule. High-reward rule trials carried an 85% probability of reward for correct responses. Low-reward rule trials carried a 15% probability of reward for correct responses. Because subjects were only rewarded for correct trials, and the lower coherence levels in the task were challenging, the effective reward rate was 64.9% for the high-reward rule and 7.6% for the low-reward rules. Subjects performed six blocks of 96 trials for a total of 572 trials of the reward task. Subjects were told that one of the blocks would be selected randomly to count for real, and rewards from that block, each worth \$0.50, would be paid as a bonus.



indicated whether subjects responded based on the shape, color, or motion of a shape kinematogram. Accurate responses on one of the rules were rewarded at a higher rate (85%) than the other two rules (15%). Feedback indicated whether the subject earned a reward on a rewarded trial (gold cross), was correct but unrewarded (no feedback), or made an incorrect response on any trial (red cross). B) After the reward period, subjects took an enforced break before commencing the test phase. This phase did not carry the possibility of reward but was otherwise identical to the reward phase. C) Each block consisted of miniblocks containing only two of the three rules. These miniblocks allowed us to compare performance on the same rule when competing against the high-reward rule versus not competing against the high-reward rule. In the above example, if motion is the rewarded rule, then color is a competing rule in a {motion, color} miniblock, and it is a non-competing rule in a {shape, color} miniblock.

To assess the effect of rule habit on cognitive flexibility, rule order was organized in an unsignaled miniblock structure. Within each miniblock, subjects performed only two out of the three rules. These miniblocks allowed us to compare performance on the same rule when competing against the high-reward rule versus not competing against the high-reward rule. Each miniblock of 16 trials contained an equal number of trials for each of the two rules in a pseudorandom order. Each miniblock, and hence the task, contained a full crossing of instructed rules and coherence levels. Each run contained six miniblocks comprising two instances of the three possible pairwise combinations of rules. Miniblocks with the same two rules were not repeated sequentially, and the first trial of each miniblock was always the rule not included in the previous miniblock. Subjects were not instructed on the miniblock structure.

Subjects were trained on the behavioral task in a two-hour session 1-3 days before the main task. Subjects first practiced each rule one at a time in blocks of 40 trials for 840 trials. During this training, the difficulty was increased by adjusting the coherence in a 3-down-one-up staircase (i.e., the coherence was reduced after three consecutive correct responses and increased after every error). Subsequently, subjects were instructed on the cue-rule assignments and performed two practice blocks of 96 trials. These blocks were identical to the main task except that they had no rewards and trivially-easy coherence. The cue-rule assignments from training were consistent for the rest of the study. Finally, subjects performed six practice blocks of the main experimental task without rewards.

### ***Data analysis***

Data were analyzed using custom code written in Python. For continuous dependent variables (e.g., reaction time), mixed-effects models were implemented using the *lmer* package in R 4.2.0 (Bates et al., 2015). For binary dependent variables (accuracy), mixed-effects models were implemented using the *glmer* package and a binomial link function. All mixed-effects models contained random intercepts for each subject and random slopes for rule coherence. We chose this random effects approach because theoretical and modeling work shows that mixed-effects models generalize most effectively when they use the maximal random effects structure that is justified by the design and does not create convergence issues (Barr et al., 2013). Data plots were created using Seaborn 0.11.2 (M. Waskom, 2021).

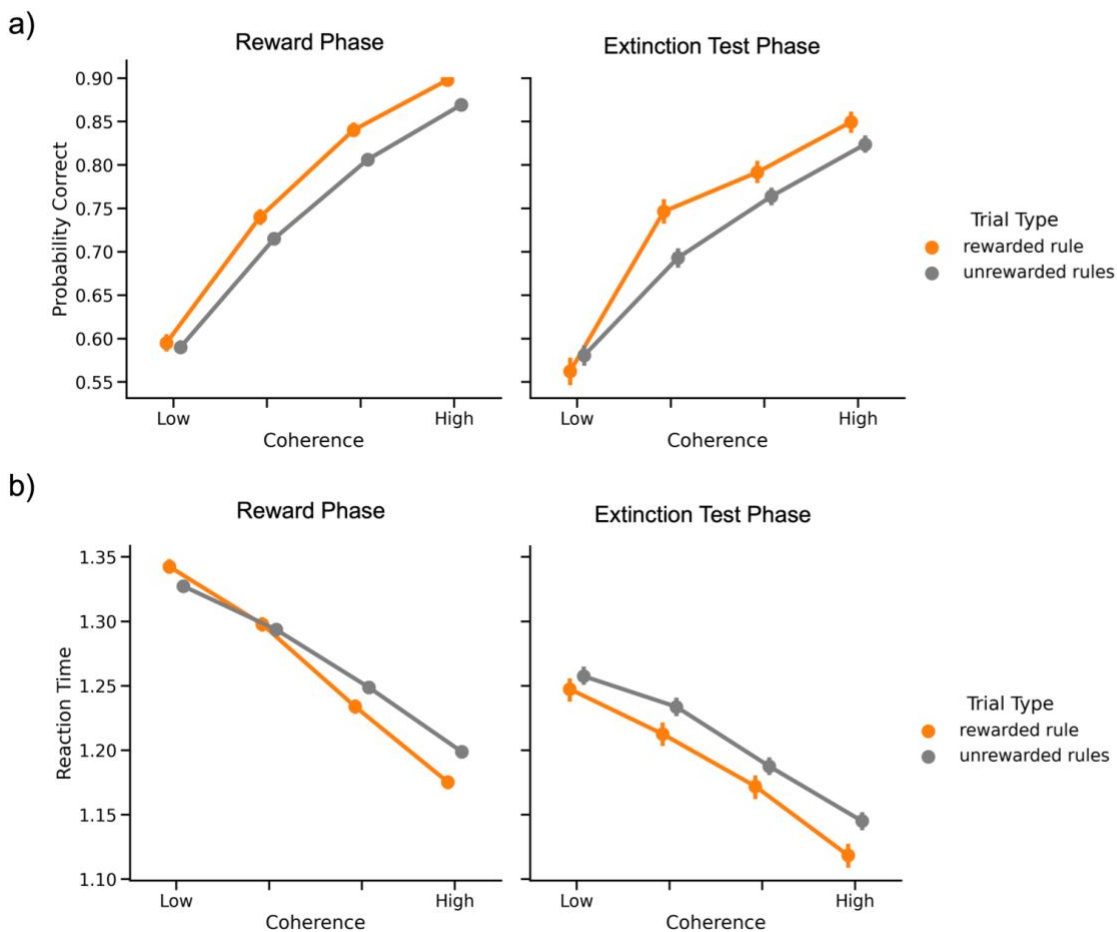
### *Drift-diffusion modeling*

Drift-diffusion modeling was performed using HDDM 9.2 (Wiecki et al., 2013). Models were fit independently for the reward reinforcement and extinction test phases using five chains of 20,000 samples. We discarded the first 10,000 samples as burn-in and thinned the chains by retaining only every fifth sample, which resulted in 10,000 samples from the posterior distribution. The Gelman-Rubin statistic was less than 1.1 ( $\max \hat{r} < 1.01$ ) for all parameters, indicating that the five chains converged to the same stationary distribution. We modeled the effects of experimental conditions on drift-rate parameters using a within-subject regression model, which allowed us to account for individual differences in task performance. Our models assumed that each subject's parameters were fixed across trials, as the more complex trial-by-trial variability models failed to converge. Models assumed a 5% outlier rate. Posterior predictive checks averaged across 500 simulations of the task for all subjects to derive predicted accuracy and reaction times.

## RESULTS

The subjects performed a rewarded, context-based perceptual discrimination task, Figure 1. One of the rules was selected as the high-reward rule, and correct performance on that rule yielded a higher reward probability than the other rules. The task is well-suited to detect habitual rule selection, as opposed to stimulus-response or feature learning, because all three stimulus features are present on every trial and have no statistical relationship to behavioral responses or rewards. We first assessed whether performance on the high-reward rule differed from the low-reward rules. Although subjects were not instructed on reward contingencies, they

were more accurate,  $z = 6.9$ ,  $p < .001$ , Figure 2A, and there was trending evidence that they were faster,  $t(4884) = 1.92$ ,  $p = .055$ , Figure 2B, when performing high-reward rules. This result is consistent with findings showing that reward motivation facilitates the execution of demanding tasks (Chiew & Braver, 2014; Krawczyk & D'Esposito, 2013; Locke & Braver, 2008). To determine whether reward reinforcement of a rule leads to the development of a rule habit, we included a post-learning test phase identical to the learning task, except that subjects were instructed that there was no possibility of reward. We predicted that implicit reward learning during the reward phase would lead to enduring facilitation of high-reward-rule execution, even when there is no longer any incentive to improve performance on the high-reward rule. During the test period, we found that accuracy was higher,  $z = 3.6$ ,  $p < .001$ , Figure 2A, and reaction times were faster,  $t(3007) = 3.92$ ,  $p < .001$ , Figure 2B, for the previously high-reward rule. This finding shows that reward reinforcement creates enduring facilitation of rule-based behavior, consistent with a cognitive habit facilitating the implementation of high-value goals.

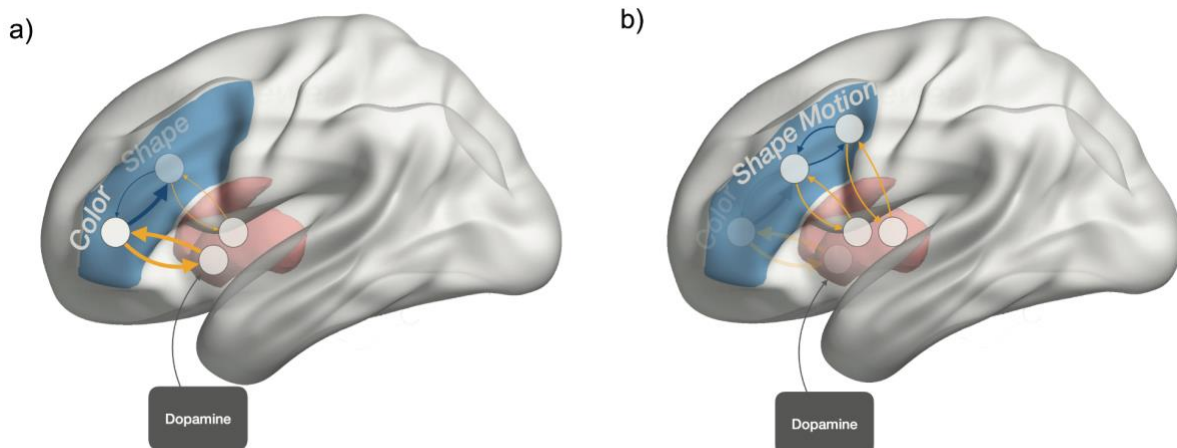




**Figure 2.** *Reward reinforcement creates rule habits.* A) Accuracy was higher for the high-reward rule than the low-reward rules in both the reward and test phases. In addition, accuracy was higher for high-coherence (easy) trials than for low-coherence (hard) trials. B) Reaction times were faster for the previously high-reward rule in the test phase. During the reward phase, there was an interaction between coherence and rule type on reaction times.

An alternative account of the improved performance of the high-reward rule is that reward biased perceptual learning of the discrimination task (i.e., the determination of the dominant color, shape, or motion direction), leading to improved perceptual discrimination in the high-reward-rule dimension (Law & Gold, 2008; Roelfsema et al., 2010; C. M. Solley & Murphy, 1960). We sought to minimize the influence of perceptual learning by training subjects on the perceptual discrimination task on a previous training day. Despite this training, there was trending evidence of continued perceptual learning during the reward phase, linear effect of trial number on accuracy,  $Z = 1.9$ ,  $p = .069$ . However, perceptual learning was not different for the high-reward rule during learning, interaction of trial number and high-reward rule  $Z = 1.5$ ,  $p = .13$ . Therefore, perceptual learning is unlikely to fully explain improved performance on the high-reward rule during the test phase.

We varied the coherence of the information in each trial to be sensitive to behavioral effects that depend on the difficulty of rule implementation (M. L. Waskom et al., 2019). As expected, coherence strongly affected rule accuracy,  $z = 24$ ,  $p < .001$ , and reaction time,  $t(4884) = -21$ ,  $p < .001$ , during the reward phase. We predicted that reward would have the largest effect on the more difficult trials because these trials benefit the most from improved rule selection and maintenance. Although we found an interaction between rule type (high-reward or low-reward) and coherence in the reward phase, accuracy  $Z = -4.3$ ,  $p < .001$ , reaction time  $t(4884) = 6.0$ ,  $p < .001$ , it was opposite to the predicted direction: we found that reward had the highest impact on easier trials. One account of this finding is that subjects stood to gain the most reward with the least cognitive effort by improving performance on easy trials (Shenhav et al., 2013). Such an adaptive cognitive control account would not predict improved performance during the test period because no rewards are at stake. Contrary to this prediction, we found a similar interaction in the test phase for rule accuracy,  $z = 2.46$ ,  $p = .014$ , though not reaction time,  $p > .2$ . In the following section, we use a drift-diffusion model approach to provide an alternative account of this behavioral effect.



**Figure 3.** *Conceptual model of rule habits.* A) Rewards reinforce corticostriatal synapses representing the rewarded rule, facilitating its selection and impairing the ability to select the non-rewarded rule. B) Selection of non-rewarded rules should be improved in contexts where rewarded rule is not relevant. The blue shaded area indicates the inferior frontal sulcus, which is preferentially engaged in context-dependent decision-making (M. L. Waskom et al., 2014), and the red shading depicts the striatum. Note that the positioning of the rules in PFC is arbitrary and is not intended to depict anterolateral gradients in feature processing.

Habits improve the execution of rewarding behaviors at the cost of reduced flexibility to adapt behavior when goals change. This facilitation is thought to arise due to dopamine-dependent adjustment of corticostriatal synapses, which facilitates selection of rules previously associated with rewards while suppressing the selection of alternative rules Figure 3A. Importantly, this reduction in flexibility should only occur when habits compete against alternative behaviors for control of behavior, Figure 3B. For example, a habit of exercising after work will specifically influence decisions about after-work plans, while not influencing decisions about the morning commute. To test for the context-dependence of the influence of goal habits on cognitive flexibility, we embedded a miniblock structure in the task, Figure 1C. Within each miniblock, subjects performed only two out of the three rules. These miniblocks created epochs where rule execution competed with a high-reward rule and epochs without this competition. During the reward phase, accuracy varied as a function of rule coherence,  $F(2.5, 210) = 709$ ,  $\eta^2g = .62$ ,  $p < .001$ , rule type (high-reward, competing, non-competing),  $F(1.4, 121) = 6.4$ ,  $\eta^2g = .016$ ,  $p = .006$ , and there was a trend towards an interaction between rule type (high-reward, competing, non-competing) and coherence,  $F(5.3, 452) = 2.1$ ,  $\eta^2g = .006$ ,  $p = .061$ , Figure 4A. These relationships persisted into the test phase, where we found a main effect of rule coherence,  $F(2.5, 210) = 215$ ,  $\eta^2g = .34$ ,  $p < .001$ , rule type (high-reward, competing, non-

competing),  $F(1.9, 157) = 3.6$ ,  $\eta^2g = .006$ ,  $p = .033$ , and an interaction between rule type and coherence,  $F(5.3, 448) = 2.8$ ,  $\eta^2g = .011$ ,  $p = .015$ . However, contrary to our predictions, there was no difference in accuracy between competing and non-competing rules in the reward nor the test phases,  $ps > .2$ . However, we found that reaction times were faster for the non-competing, relative to the competing rule during the reward phase,  $t(4884) = 6.16$ ,  $p < .001$ , Figure 4C, but not during the test phase,  $p > .2$ . Given that this analysis compares performance on the same rules in different miniblock contexts, we may have lacked the sensitivity to detect small differences in performance.

As an alternative method of testing whether reward created a rule habit, we assessed whether switch costs in reaction time varied as a function of reward. Because a habit should facilitate the selection of high-reward rules, we predicted that switching *to* a high-reward rule should be faster than switching between low-reward rules. Additionally, because habits can impair the flexibility to adapt behavior when goals change, we predicted that switching *away* from a high-reward rule ought to be slower than switching between low-reward rules. We constructed a model with switch type {switching away from a high-reward rule, switching to a high reward rule, staying with the same rule, and switching between low reward rules} and trial-type {high-reward, competing, non-competing} as regressors. The trial-type regressor ensures that any differential switch costs are not simply due to performance differences between the rules. Consistent with the predictions of the rule habit hypothesis, we found that during the reward phase, subjects were faster at switching to the high-reward rule, relative to switching between low-reward rules,  $Z = -5.3$ ,  $p < .001$ . However, switching away from a high-reward rule was not slower than switching between low-reward rules,  $p > .2$ . These results suggest that the selection of a high-reward rule is facilitated during reward learning, while the flexibility to adapt behavior is unaffected by reward. In contrast, in the test phase, we found that subjects were slower to switch away from a high-reward rule, relative to switching between low-reward rules,  $Z = 1.96$ ,  $p = .049$ . Consistent with habitual goal selection, this finding shows persistently reduced flexibility to adapt behavior away from previously high-reward rules.

### ***Drift diffusion models of choice***

*Reward influences multiple components of rule-guided behavior*

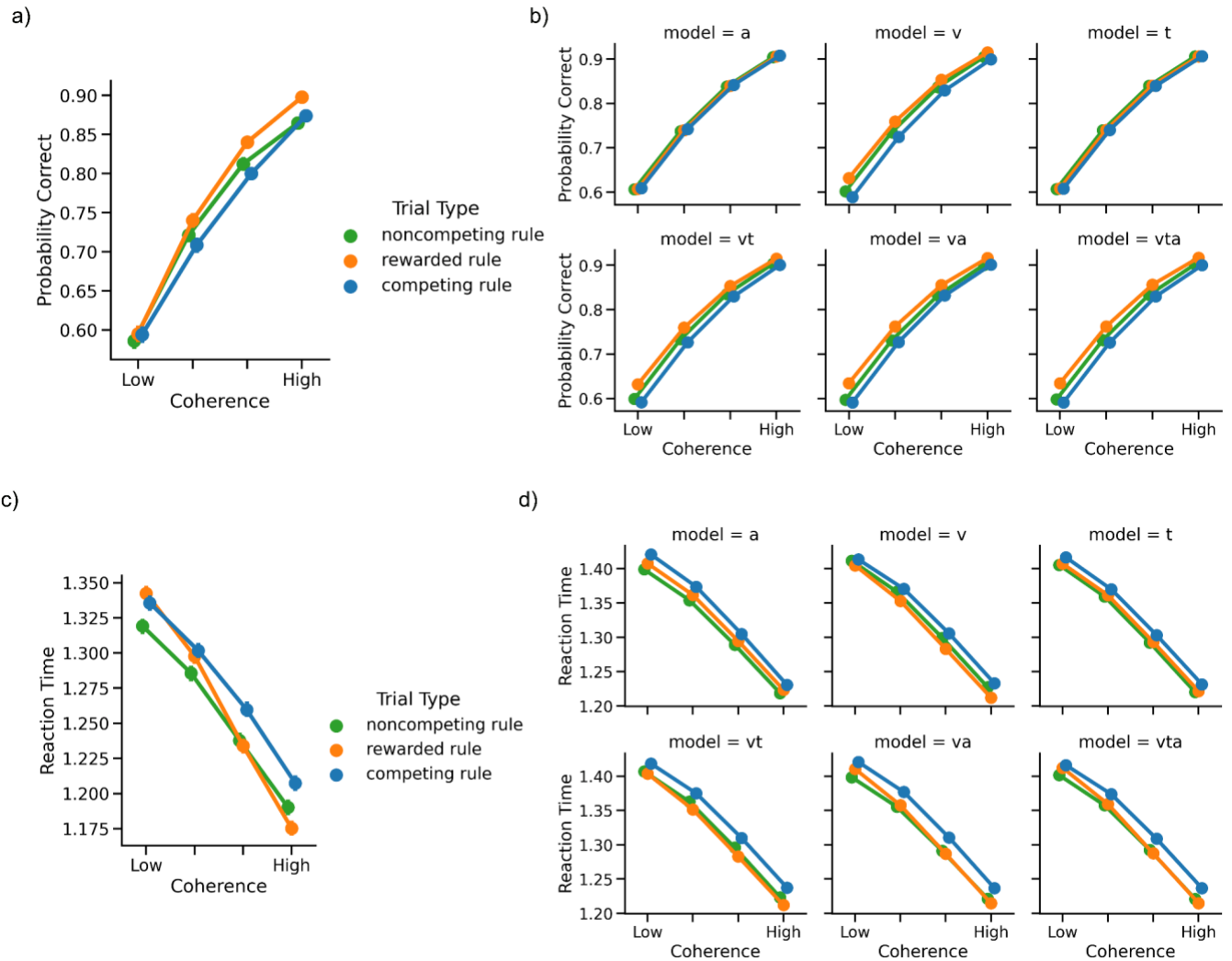
We predicted that reward reinforcement would lead to the formation of a rule habit that both facilitates the execution of the habitized rule and impairs the ability to implement competing rules. Reward could influence performance on rule behavior in multiple different ways. For example, reward could facilitate the initial selection of the rule while not influencing the execution of the rule. Alternatively, reward could influence the execution of the rule by influencing the fidelity of sensory representations of the relevant stimulus dimension (Goltstein et al., 2018; Hickey et al., 2015) or by shifting the speed-accuracy tradeoff in favor of accuracy (Bogacz et al., 2006; Tajima et al., 2016). We fit a drift-diffusion model to our data to discover how reward influences rule-guided behavior. Drift-diffusion models (DDMs) conceptualize decision-making as an evidence-accumulation process that commits to a decision when the threshold of evidence for an option is crossed. This framework deconstructs complex rule-guided behavior into distinct behavioral components, which allows for precise hypothesis testing about the influence of reward and the formation of high-level habits.

We sought to establish whether rule type (i.e., rewarded, competing, noncompeting) influenced three independent aspects of the decision-making process:

1) The *drift rate* captures the efficiency of the evidence integration process. Conditions with higher drift rates will have higher accuracy and faster reaction times. This parameter can capture variability in rule execution between rule types.

2) The *decision threshold* captures the level of evidence needed to commit to a decision. Conditions with higher thresholds will have higher accuracy and slower reaction times. This parameter can capture variability in response caution between rule types.

3) The *non-decision time* captures the time needed to initiate the drift-diffusion process. Conditions with higher non-decision time will have slower reaction times without necessarily having higher accuracy. This parameter can capture variability in the time it takes to select a rule.

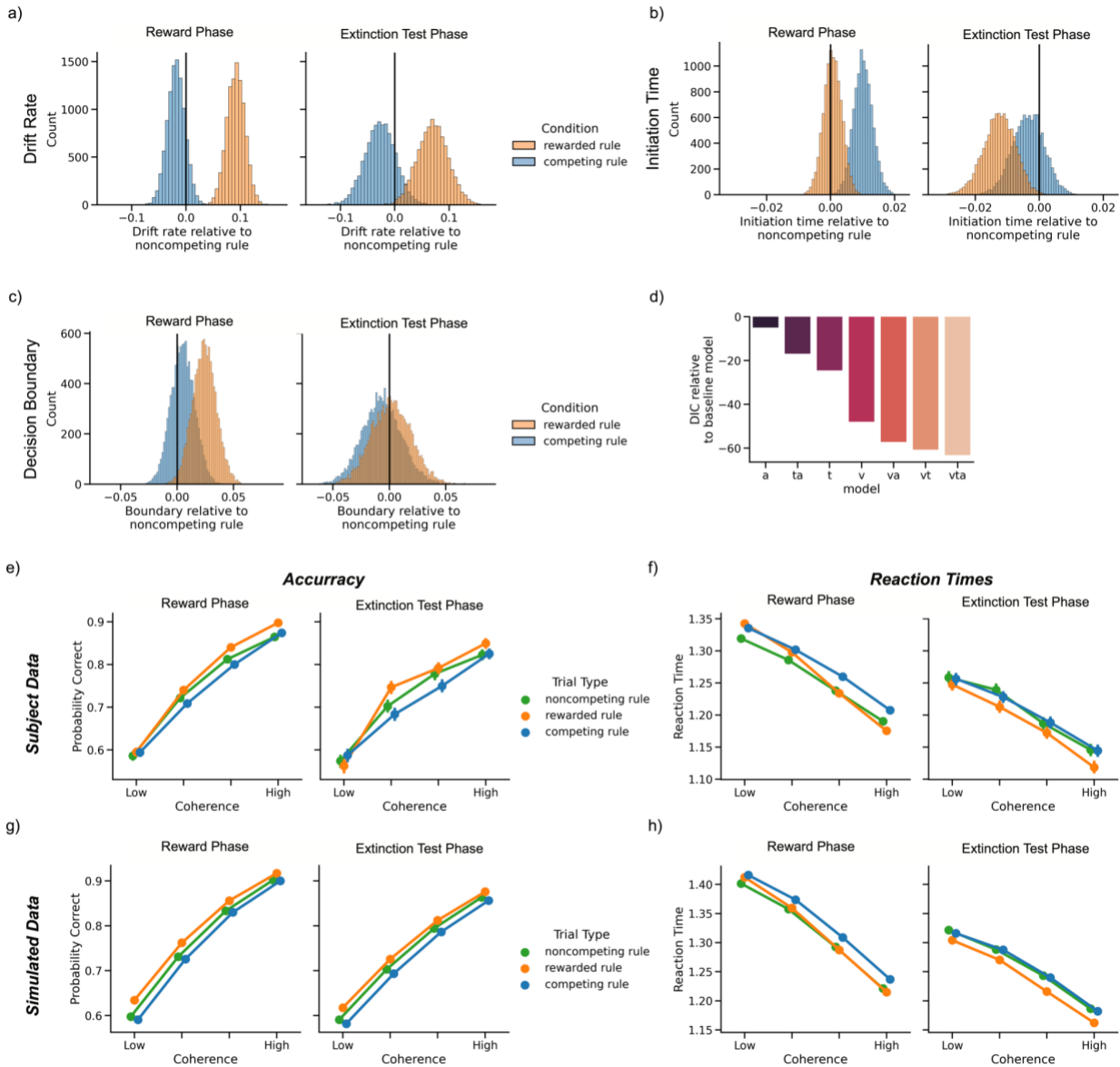


**Figure 4. Qualitative model comparison.** Models are labeled by parameters that were influenced by rule condition:  $a$  = decision threshold,  $t$  = initiation time, and  $v$  = drift rate. A) Subject accuracy by rule coherence and experimental condition for the reward phase. Subjects are more accurate for the rewarded rule. B) Simulated accuracy data for each model. Only models in which the task condition influenced the drift rate could explain the increased accuracy for the high-reward rule. C) Subject reaction times by rule coherence and experimental condition for the reward phase. Subjects are faster for the noncompeting relative to the competing rule. In addition, reaction times for the rewarded rule are faster on easy relative to hard trials. D) Simulated reaction time data for each model. Most models could capture the reaction time difference between competing and non-competing rules. However, only the models in which both drift rate and decision threshold are influenced by reward could capture the interaction between reward condition and coherence on reaction time.

We first sought to establish whether rule type influenced each of these parameters. Our model selection strategy employed a balanced consideration of both formal model quantification, the *deviance information criterion* (DIC), and post-predictive checks, which compare models by asking whether they explain qualitative features of interest in the data

(Wilson & Collins, 2019). We wanted our models to explain three prominent effects in the reward phase data: (1) higher accuracy for the high-reward rule and, Figure 4A (2) the interaction between coherence and rule type on accuracy and reaction time, Figure 4C, and (3) the slower reaction time for the competing, relative to the noncompeting rule, Figure 4C. We first examined three models in which rule type (high-reward, competing, or noncompeting), influenced drift rate, decision threshold, or non-decision time only. Of these models, only the drift-rate model could accurately capture the higher accuracy for the high-reward rule trials, Figure 4B. However, this model failed to capture the other two behavioral effects, Figure 4D. We next asked whether including additional effects of rule type on decision threshold or non-decision time could capture these effects. We found that both the {drift-rate, threshold} models and the full {drift-rate, threshold, non-decision time} models could capture all three qualitative behavioral effects, Figure 4B,D. We decided to use the {drift-rate, threshold, non-decision time} model as the final model for our data for three reasons:

1. The post-predictive checks show a small but significant improvement in the model's ability to account for the data, Figure 4D.
2. It had the lowest DIC, Figure 5D, indicating that it provided the best balance between explanatory power and complexity of the models we compared.
3. An effect of the task condition on each of these three parameters has distinct interpretations, and the full model avoids the risk of misinterpreting results (e.g., by attributing an effect to the decision threshold that would have been better captured by non-decision time).



**Figure 5.** Reward reinforcement creates a persistent increase in cognitive efficiency at the cost of reduced cognitive flexibility. A) Drift rates were higher for the high-reward rule during both reward reinforcement and the test phase, indicating persistent enhancement of rule implementation by reward. Moreover, drift rates were moderately reduced for the competing rule relative to the non-competing rule in both phases, indicating by the leftward shift of the competing rule distribution relative to  $x = 0$ , the non-competing rule reference B) Initiation times, which reflect the time it takes to begin evidence integration, were faster for the high-reward rule relative to the competing rule in both task blocks C) In the reward phase, decision thresholds were increased for the high-reward rule, indicating increased response caution. This strategy optimizes the reward rate by favoring slower, more accurate responses. D) Bayesian model comparison favors a model in which rule condition influences drift rate, initiation times, and decision thresholds. Lower DIC scores indicate more model evidence, and scores are defined relative to a baseline model without condition effects. Models are labeled by parameters that

were influenced by rule condition:  $a$  = decision threshold,  $t$  = initiation time, and  $v$  = drift rate. E) Subject accuracy and F) reaction times for each rule condition are well-matched by simulated data (G-H) from the model.

### *Impacts of reward on rule execution*

We theorized that reward reinforcement creates habits that facilitate rule execution. The drift rate parameter of a DDM reflects the sensitivity of the evidence integration process, with higher drift rates corresponding to improved rule execution. We first asked whether drift rates were increased for the high-reward rule. Unlike null hypothesis significance testing, which tries to reject a null hypothesis, Bayesian posterior probabilities indicate the model's evidence that an effect exists, given the data. We considered *strong evidence* to be when more than 95% of the posterior distribution was larger in one condition than another, *moderate evidence* to be 90% or higher, and *trending evidence* to be 80% or higher. We found that the drift rates were higher for the high-reward rule relative to the non-competing rule, *posterior probability* = 100%, and the competing rule, *posterior probability* = 100%. Consistent with the development of a rule habit, we found that this facilitation of the high-reward rule persisted into the test phase, *posterior probability* of reward > noncompete = 99.6%, reward > compete = 100%, Figure 5a. Because habits can impair the flexibility to change behavior, we next examined whether execution of rules that compete against high-reward rules was impaired. We found trending evidence that drift rates were lower for competing rules than non-competing rules in both the reward phase, *posterior probability* = 86.9%, and the test phase, *posterior probability* = 83.3%, Figure 5a. Together, these findings suggest that reward reinforcement creates a lasting improvement in rule execution, potentially at the cost of reduced flexibility to implement alternative, less-rewarding rules.

An important question posed by our findings is how reward reinforcement improves execution of the high-reward rule. One possibility is that strengthening the rule representation reduces interference from alternative rules. Dopamine release in PFC may help maintain the current rule in working memory and gate information from the irrelevant stimulus dimensions (Cools & D'Esposito, 2011; O'Reilly & Frank, 2006). According to this view, drift rates are lower for the low-reward rules because subjects are less adept at filtering irrelevant stimulus information. We asked whether response information from the other dimensions influenced behavior (e.g., the motion direction on a color rule trial). Consistent with reward protecting rules from interference, reaction times were relatively faster for the high-reward rule when the dimensions of the kinematogram indicated conflicting button responses,  $t(48890) = -2.0$ ,  $p =$



.048. This result suggests that reward improves rule execution by reducing interference from lower reward rules.

### *Impacts of reward on rule selection*

We theorized that a habit would facilitate the selection of the high-reward rule because dopaminergic modulation of corticostriatal circuitry ought to facilitate gating of high-reward representations (O'Reilly & Frank, 2006). Variation in the initiation time parameter of the DDM provides a proxy for the time it takes to internally select a rule because rule selection likely occurs prior to rule execution in this task. Consistent with the habit model, we found that the initiation time was reduced for the high-reward rule relative to the competing rule during the test phase, *posterior probability* = 99.8%, though not during the reward phase (Leong et al., 2019), Figure 5b. It is possible that the development of faster rule selection occurs slowly and is only detectable after training. We also predicted that habits would interfere with the selection of alternative behaviors. We tested whether subjects were slower to initiate low-reward rules that were competing in a context with high-reward rules. During the reward phase, we found that the initiation time of the competing rule was increased relative to the non-competing rule, *posterior probability* = 100%, Figure 5b. This result suggests that reward reduces the flexibility to select competing, nonreward rules. However, during the test phase, the initiation time of the competing rule was not slower than the non-competing rule, *posterior probability* = 75.2%, Figure 5b. This finding suggests that competing rules are harder to select when rewards are available, but this reduced flexibility does not persist after reward learning.

### *Impacts of reward on response caution*

Although our model selection indicated that rule type influenced decision threshold, we did not predict any influence of habit on response caution. However, during the reward phase, subjects were incentivized to respond accurately. We predicted that reward reinforcement would increase the decision thresholds for the high-reward rule because this strategy optimizes reward in a context where accuracy is more important than reaction time (Bogacz et al., 2006; Tajima et al., 2016). Consistent with this prediction, we found that the decision thresholds were higher for the high-reward rule relative to both the non-competing rule, *posterior probability* = 98.8%, and the competing rule, *posterior probability* = 95.7%, Figure 5c. Because this adjustment of the decision threshold is adaptive for earning rewards, we did not expect it to persist into the test phase. Rule condition did not influence decision thresholds in the test phase, *posterior*

probabilities < 70%, Figure 5c. Additionally, there were no differences between the competing and non-competing rules in either task phase, posterior probabilities < 70%. In sum, subjects adjusted their response caution adaptively, responding more carefully *only* on trials where rewards were likely (Grahek et al., 2021). This suggests that different components of rule-guided decision-making are differentially sensitive to the effects of reward reinforcement, with persistent biases emerging in rule execution (drift rate) and rule selection (initiation time) but not in the selection of an appropriate speed-accuracy tradeoff (decision threshold). We speculate that this distinction occurs because setting a decision threshold is a superordinate control process for determining a decision strategy and may be more sensitive to changing reward values (Frank, 2006; Son & Sethi, 2006).

Because our DDM disentangles the effects of rule type on distinct components of rule-guided behavior, it can provide a mechanistic account of surprising behavioral effects. We observed that reward led to faster reaction times only on easier trials, Figure 5f, left panel. The DDM shows that this effect arises because reward influences both drift-rates and decision thresholds. Increased drift rates lead to overall faster reaction times for high-reward rules. However, increased decision thresholds cause slower reaction times for harder, high-reward rules. Together, these factors predict that the reward will cause the fastest reaction times on easy trials, the effect found in our data, Figure 5h. However, the model also predicts improved accuracy on the most difficult high-reward trials, an effect that is not present in our data. The most difficult trials may engage maximal cognitive resources (Kool et al., 2017), and reward may have no additive benefit above and beyond intrinsic motivation. Future work is needed to disentangle the influence of extrinsic and intrinsic motivation on cognitive control (Dobryakova et al., 2017; Sullivan-Toole et al., 2017).

## DISCUSSION

We found evidence that reward reinforcement leads to the development of habitual rule selection that persists in the absence of reward. Our findings are consistent with the theory that dopamine release adjusts corticostriatal synaptic plasticity to favor the selection of rewarding rules. However, there are several potential mechanisms by which reward could influence the performance of the high-reward rule. We argue that abstract rule representations in PFC are reinforced by reward, which facilitates activation and implementation of the rule. Our findings that initiation times are reduced and drift rates are enhanced for the high-reward rule are

consistent with this finding. A related mechanism is that subjects learn to attribute value to the rule cues (i.e., the central shapes, Figure 1) rather than internal rule representations. According to this model, valuable cues trigger motivation to use cognitive control without being linked to a specific rule (Ballard et al., 2011; Niv et al., 2006; Shenhav et al., 2013). This model could explain improved performance on the high-reward rule but does not explain the finding that subjects are impaired at switching away from high-reward rules. In fact, increased motivation should increase cognitive flexibility (Aarts et al., 2010; Dreisbach, 2006). Another potential mechanism is that reward strengthens the associative link between the cue and its associated rule representation (Miller et al., 2002). This strengthened association could facilitate activation of the rule representation, which would account for the finding that subjects showed reduced initiation times for the high-reward rule. However, this model does not explain why drift rates are increased for the high-reward rule. Nonetheless, these different mechanisms may co-occur, and future work is warranted to identify how reward reinforces internal cue and goal representations.

Rewards likely influenced rule execution by biasing the allocation of attention to the the rule-relevant dimension (Etzel et al., 2016; M. L. Waskom et al., 2014). We posit that reward influences the deployment of top-down attention to facilitate sensory evidence integration (Botvinick & Braver, 2015; Frömer et al., 2021; Krebs & Woldorff, 2017). However, it is also possible that reward increases the salience of the high-reward-dimension features (e.g., color), which captures attention in a bottom-up manner (Anderson et al., 2011; Della Libera & Chelazzi, 2009; Failing & Theeuwes, 2014). One distinction between the top-down and bottom-up accounts is that the bottom-up mechanism does not predict any difference in performance between the competing and non-competing rules (because these are the same rules in different contexts). However, evidence for a difference in performance between competing versus non-competing rules was mixed. Although reaction times differed between these conditions during training, this difference did not persist into the test phase. Further, the drift-diffusion models found only moderate evidence for a difference in drift rate between these conditions. Therefore, it is plausible that both top-down and bottom-up attentional mechanisms could contribute to improved performance on the high-reward rule (Grahek et al., 2021). Importantly, both top-down and bottom-up attentional mechanisms likely contribute to habitual goal selection. For example, images of cigarettes in the media can capture the attention of smokers, which could then activate the goal of purchasing cigarettes (Versace et al., 2010; Wood & Runger, 2016).

We found that initiation times were faster for the non-competing relative to the competing rule during the reward phase of the experiment. We argue that this finding demonstrates reduced flexibility to adapt behavior away from a habitized rule. An alternative interpretation is that participants allocate cognitive effort according to the relative value of each rule within a miniblock. This scaling of reward value relative to the context is termed *range adaption* (Hunter & Daw, 2021; Tversky & Kahneman, 1986). In our task, the higher overall reward rate during the reward rule miniblocks could render the small reward possibility associated with the competing rule comparatively less valuable. In the non-competing miniblocks, the lower reward rate renders the small reward probability of the non-competing rules relatively more valuable. This account is consistent with our finding that there was no difference in initiation time between the competing and non-competing rules in the test phase, when the reward rates of the two conditions were matched. However, we also found that switch costs were higher when dropping the previously high-reward rule in the test phase. This finding is inconsistent with range adaptation because there are no rewards in the test phase. Moreover, we found similar, though moderate, differences in drift rates between competing and non-competing rules in both the reward and test phases. Nonetheless, the question of how range adaptation interacts with the development of a goal habit is an important target for future research.

A key feature of habits is that they are difficult to unlearn. Our measures showed persistent biases in a test period without rewards that occurred several minutes after reward learning, indicating resistance to new learning. However, the experiment was not designed to create or assess long-lasting habits. The long-term resiliency of habits likely relies on additional neural mechanisms, such as increasing dorsal striatal involvement in decision-making (Yin & Knowlton, 2006). Moreover, durable habits are extremely difficult to elicit in the laboratory due to the flexibility of human goal-directed cognition (de Wit et al., 2018; Hardwick et al., 2019). Future work is needed to explore the interacting psychological conditions, including temporally extended learning and reward anticipation (Ballard et al., 2017; Yin & Knowlton, 2006), as well as factors such as stress (Schwabe & Wolf, 2009) and social motivation (Wood, 2017), underlying the development of long-lasting habitual goal selection.

Our results show that reward reinforcement creates persistent facilitation of rule-guided behavior at the expense of reduced cognitive flexibility. However, significant differences exist between the constructs in our task and real-world goal habits. A critical question for psychological research concerns the nature of reinforcement in ecological situations (Brewer &

Roy, 2021). Primary reinforcers, such as money or food, likely act alongside abstract reinforcers, such as goal attainment (McDougle et al., 2021; Swanson & Tricomi, 2014) in forming cognitive habits. The habit system is involved in a variety of disorders, including anxiety (Lago et al., 2017; Packard, 2009), OCD (Gillan et al., 2014; Gillan & Robbins, 2014), anorexia nervosa (Foerde et al., 2021; Steinglass & Walsh, 2006) and Parkinson's disease (Cools, 2011; Dubois & Pillon, 1997). Understanding the role of cognitive habits in these disorders will require an account of whether and why the dopaminergic system reinforces maladaptive goals. Our study focused on the idea that a goal habit involves the habitual selection of a goal, which is then flexibility implemented (Cushman & Morris, 2015). Future work is needed to understand whether pursuing habitually selected goals has a temporally-extended influence on cognitive flexibility. An improved understanding of the etiology of cognitive habits may lead to novel strategies for acquiring adaptive habits and weakening habits that impair cognitive well-being.

## DATA AVAILABILITY STATEMENT

All code and behavioral data will be available on GitHub prior to publication.

## AUTHOR CONTRIBUTIONS

ICB: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing- Original Draft, Visualization. MW: Conceptualization, Methodology, Writing- Review and Editing. KN: Methodology, Software, Formal Analysis, Investigation, Writing- Review and Editing. MD: Conceptualization, Methodology, Supervision, Funding Acquisition, Writing- Review and Editing.

## ACKNOWLEDGEMENTS

We thank the D'Esposito lab and Cognitive Computational Neuroscience lab (PI: Anne Collins) for helpful discussions. We would also like to thank Debbie Yee, Ivan Grahek, and the HDDM online community for their advice on the HDDM package.

## FUNDING INFORMATION

ICB is funded by the NIH fellowship F32MH119796. The research was supported by the National Institute of Health grant MH063901.

## REFERENCES

- Aarts, E., Roelofs, A., Franke, B., Rijpkema, M., Fernández, G., Helmich, R. C., & Cools, R. (2010). Striatal dopamine mediates the interface between motivational and cognitive control in humans: evidence from genetic imaging. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *35*(9), 1943–1951. <https://doi.org/10.1038/npp.2010.68>
- Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, *9*, 357–381. <https://doi.org/10.1146/annurev.ne.09.030186.002041>
- Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(25), 10367–10371. <https://doi.org/10.1073/pnas.1104047108>
- Badre, D., & Frank, M. J. (2012). Mechanisms of Hierarchical Reinforcement Learning in Cortico–Striatal Circuits 2: Evidence from fMRI. *Cerebral Cortex*, *22*(3), 527–536. <https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhr117>
- Ballard, I. C., Hennigan, K., & McClure, S. M. (2017). Mere exposure: Preference change for novel drinks reflected in human ventral tegmental area. *Journal of Cognitive Neuroscience*. [https://www.mitpressjournals.org/doi/abs/10.1162/jocn\\_a\\_01098](https://www.mitpressjournals.org/doi/abs/10.1162/jocn_a_01098)
- Ballard, I. C., Murty, V. P., Carter, R. M., MacInnes, J. J., Huettel, S. A., & Adcock, R. A. (2011). Dorsolateral prefrontal cortex drives mesolimbic dopaminergic regions to initiate motivated behavior. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(28), 10340–10346. <https://doi.org/10.1523/JNEUROSCI.0895-11.2011>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. <https://doi.org/10.18637/jss.v067.i01>

- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765. <https://doi.org/10.1037/0033-295X.113.4.700>
- Botvinick, M., & Braver, T. (2015). Motivation and cognitive control: from behavior to neural mechanism. *Annual Review of Psychology*, *66*, 83–113. <https://doi.org/10.1146/annurev-psych-010814-015044>
- Brewer, J. A., & Roy, A. (2021). Can Approaching Anxiety Like a Habit Lead to Novel Treatments? *American Journal of Lifestyle Medicine*, *15*(5), 489–494. <https://doi.org/10.1177/15598276211008144>
- 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. <https://doi.org/10.3758/s13415-014-0280-0>
- Collins, A. G. E., & Frank, M. J. (2013). Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychological Review*, *120*(1), 190–229. <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=23356780&retmode=ref&cmd=prlinks>
- Cools, R. (2011). Dopaminergic control of the striatum for high-level cognition. *Current Opinion in Neurobiology*, *21*(3), 402–407. <https://doi.org/10.1016/j.conb.2011.04.002>
- Cools, R., & D'Esposito, M. (2011). Inverted-U-Shaped Dopamine Actions on Human Working Memory and Cognitive Control. *Biological Psychiatry*, *69*(12), e113–e125. <http://linkinghub.elsevier.com/retrieve/pii/S0006322311002782>
- Cushman, F., & Morris, A. (2015). Habitual control of goal selection in humans. *Proceedings of the National Academy of Sciences*, *112*(45), 13817–13822. <https://doi.org/10.1073/pnas.1506367112>
- Della Libera, C., & Chelazzi, L. (2009). Learning to attend and to ignore is a matter of gains and

- losses. *Psychological Science*, 20(6), 778–784. <https://doi.org/10.1111/j.1467-9280.2009.02360.x>
- DeLong, M. R. (1990). Primate models of movement disorders of basal ganglia origin. *Trends in Neurosciences*, 13(7), 281–285. [https://doi.org/10.1016/0166-2236\(90\)90110-v](https://doi.org/10.1016/0166-2236(90)90110-v)
- de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., Evans, M., Mirza, H., & Gillan, C. M. (2018). Shifting the balance between goals and habits: Five failures in experimental habit induction. *Journal of Experimental Psychology. General*, 147(7), 1043–1065. <https://doi.org/10.1037/xge0000402>
- Dobryakova, E., Jessup, R. K., & Tricomi, E. (2017). Modulation of ventral striatal activity by cognitive effort. *NeuroImage*, 147, 330–338. <https://doi.org/10.1016/j.neuroimage.2016.12.029>
- Dreisbach, G. (2006). How positive affect modulates cognitive control: the costs and benefits of reduced maintenance capability. *Brain and Cognition*, 60(1), 11–19. <https://doi.org/10.1016/j.bandc.2005.08.003>
- Dubois, B., & Pillon, B. (1997). Cognitive deficits in Parkinson's disease. *Journal of Neurology*, 244(1), 2–8. <https://doi.org/10.1007/pl00007725>
- Dworkin, J. D., Linn, K. A., Teich, E. G., Zurn, P., Shinohara, R. T., & Bassett, D. S. (2020). The extent and drivers of gender imbalance in neuroscience reference lists. *Nature Neuroscience*, 23(8), 918–926. <https://doi.org/10.1038/s41593-020-0658-y>
- Eckstein, M. K., & Collins, A. G. E. (2020). Computational evidence for hierarchically structured reinforcement learning in humans. *Proceedings of the National Academy of Sciences*, 117(47), 29381–29389. <https://doi.org/10.1073/pnas.1912330117>
- Etzel, J. A., Cole, M. W., Zacks, J. M., Kay, K. N., & Braver, T. S. (2016). Reward Motivation Enhances Task Coding in Frontoparietal Cortex. *Cerebral Cortex*, 26(4), 1647–1659. <https://doi.org/10.1093/cercor/bhu327>
- Failing, M. F., & Theeuwes, J. (2014). Exogenous visual orienting by reward. *Journal of Vision*,



14(5), 6. <https://doi.org/10.1167/14.5.6>

- Foerde, K., Walsh, B. T., Dalack, M., Daw, N., Shohamy, D., & Steinglass, J. E. (2021). Changes in brain and behavior during food-based decision-making following treatment of anorexia nervosa. *Journal of Eating Disorders*, 9(1), 48. <https://doi.org/10.1186/s40337-021-00402-y>
- Frank, M. J. (2006). Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks: The Official Journal of the International Neural Network Society*, 19(8), 1120–1136. <https://doi.org/10.1016/j.neunet.2006.03.006>
- Frank, M. J., & Badre, D. (2012). Mechanisms of Hierarchical Reinforcement Learning in Corticostriatal Circuits 1: Computational Analysis. *Cerebral Cortex*, 22(3), 509–526. <https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhr114>
- 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), 1–11. <https://doi.org/10.1038/s41467-021-21315-z>
- Fulvio, J. M., Akinnola, I., & Postle, B. R. (2021). Gender (Im)balance in Citation Practices in Cognitive Neuroscience. *Journal of Cognitive Neuroscience*, 33(1), 3–7. [https://doi.org/10.1162/jocn\\_a\\_01643](https://doi.org/10.1162/jocn_a_01643)
- Gillan, C. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Voon, V., Apergis-Schoute, A. M., Fineberg, N. A., Sahakian, B. J., & Robbins, T. W. (2014). Enhanced avoidance habits in obsessive-compulsive disorder. *Biological Psychiatry*, 75(8), 631–638. <https://doi.org/10.1016/j.biopsych.2013.02.002>
- Gillan, C. M., & Robbins, T. W. (2014). Goal-directed learning and obsessive-compulsive disorder. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1655). <https://doi.org/10.1098/rstb.2013.0475>
- Goltstein, P. M., Meijer, G. T., & Pennartz, C. M. A. (2018). Conditioning sharpens the spatial representation of rewarded stimuli in mouse primary visual cortex. *eLife*, 7, e37683.

<https://doi.org/10.7554/eLife.37683>

Grahek, I., Schettino, A., Koster, E. H. W., & Andersen, S. K. (2021). Dynamic Interplay between Reward and Voluntary Attention Determines Stimulus Processing in Visual Cortex. *Journal of Cognitive Neuroscience*, 33(11), 2357–2371.

[https://doi.org/10.1162/jocn\\_a\\_01762](https://doi.org/10.1162/jocn_a_01762)

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory*, 70(1-2), 119–136.

<http://linkinghub.elsevier.com/retrieve/pii/S1074742798938436>

Haber, S. N. (2012). Neuroanatomy of Reward: A View from the Ventral Striatum. In J. A. Gottfried (Ed.), *Neurobiology of Sensation and Reward*. CRC Press/Taylor & Francis.

<https://www.ncbi.nlm.nih.gov/pubmed/22593898>

Hardwick, R. M., Forrence, A. D., Krakauer, J. W., & Haith, A. M. (2019). Time-dependent competition between goal-directed and habitual response preparation. *Nature Human Behaviour*, 3(12), 1252–1262. <https://doi.org/10.1038/s41562-019-0725-0>

Hickey, C., Kaiser, D., & Peelen, M. V. (2015). Reward guides attention to object categories in real-world scenes. *Journal of Experimental Psychology. General*, 144(2), 264–273.

<https://doi.org/10.1037/a0038627>

Hunter, L. E., & Daw, N. D. (2021). Context-sensitive valuation and learning. *Current Opinion in Behavioral Sciences*, 41, 122–127. <https://doi.org/10.1016/j.cobeha.2021.05.001>

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans - ProQuest. *Science*.

<http://search.proquest.com/openview/16908e7d27610a0e04267a4b47d48b12/1?pq-origsite=gscholar&cbl=1256>

Kool, W., Shenhav, A., & Botvinick, M. M. (2017). Cognitive control as cost-benefit decision making. In *The Wiley Handbook of Cognitive Control* (pp. 167–189). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118920497.ch10>

- Krawczyk, D. C., & D'Esposito, M. (2013). Modulation of working memory function by motivation through loss-aversion. *Human Brain Mapping, 34*(4), 762–774.  
<https://doi.org/10.1002/hbm.21472>
- Krebs, R. M., & Woldorff, M. G. (2017). Cognitive control and reward. In T. Egner (Ed.), *The Wiley handbook of cognitive control*, (pp (Vol. 642, pp. 422–439).  
<https://doi.org/10.1002/9781118920497.ch24>
- Lago, T., Davis, A., Grillon, C., & Ernst, M. (2017). Striatum on the anxiety map: Small detours into adolescence. *Brain Research, 1654*(Pt B), 177–184.  
<https://doi.org/10.1016/j.brainres.2016.06.006>
- Law, C.-T., & Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nature Neuroscience, 11*(4), 505–513.  
<https://doi.org/10.1038/nn2070>
- Leong, Y. C., Hughes, B. L., Wang, Y., & Zaki, J. (2019). Neurocomputational mechanisms underlying motivated seeing. *Nature Human Behaviour, 3*(9), 962–973.  
<https://doi.org/10.1038/s41562-019-0637-z>
- Lhermitte, F. (1983). “Utilization behaviour” and its relation to lesions of the frontal lobes. *Brain: A Journal of Neurology, 106*(2), 237–255. <https://academic.oup.com/brain/article-abstract/106/2/237/299596>
- Locke, H. S., & Braver, T. S. (2008). Motivational influences on cognitive control: behavior, brain activation, and individual differences. *Cognitive, Affective & Behavioral Neuroscience, 8*(1), 99–112. <https://doi.org/10.3758/cabn.8.1.99>
- McDougle, S. D., Ballard, I. C., Baribault, B., Bishop, S. J., & Collins, A. G. E. (2021). Executive Function Assigns Value to Novel Goal-Congruent Outcomes. *Cerebral Cortex, 32*(1), 231–247. <https://doi.org/10.1093/cercor/bhab205>
- Miller, E. K., Freedman, D. J., & Wallis, J. D. (2002). The prefrontal cortex: categories, concepts and cognition. *Philosophical Transactions of the Royal Society of London. Series B,*

- Biological Sciences*, 357(1424), 1123–1136. <https://doi.org/10.1098/rstb.2002.1099>
- Murty, V. P., Tompary, A., Adcock, R. A., & Davachi, L. (2017). Selectivity in Postencoding Connectivity with High-Level Visual Cortex Is Associated with Reward-Motivated Memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 37(3), 537–545. <https://doi.org/10.1523/JNEUROSCI.4032-15.2016>
- Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3), 139–154. <http://linkinghub.elsevier.com/retrieve/pii/S0022249608001181>
- Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2006). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, 191(3), 507–520. <http://www.springerlink.com/index/10.1007/s00213-006-0502-4>
- O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, 18(2), 283–328. <http://www.mitpressjournals.org/doi/10.1162/089976606775093909>
- Packard, M. G. (2009). Anxiety, cognition, and habit: a multiple memory systems perspective. *Brain Research*, 1293, 121–128. <https://doi.org/10.1016/j.brainres.2009.03.029>
- Radulescu, A., Niv, Y., & Ballard, I. (2019). Holistic Reinforcement Learning: The Role of Structure and Attention. *Trends in Cognitive Sciences*, 23(4), 278–292. <https://doi.org/10.1016/j.tics.2019.01.010>
- Ribas-Fernandes, J. J. F., Solway, A., Diuk, C., McGuire, J. T., Barto, A. G., Niv, Y., & Botvinick, M. M. (2011). A Neural Signature of Hierarchical Reinforcement Learning. *Neuron*, 71(2), 370–379. <http://linkinghub.elsevier.com/retrieve/pii/S0896627311004995>
- Rmus, M., McDougale, S. D., & Collins, A. G. E. (2021). The Role of Executive Function in Shaping Reinforcement Learning. *Current Opinion in Behavioral Sciences*, 38, 66–73. <https://doi.org/10.1016/j.cobeha.2020.10.003>
- Roelfsema, P. R., van Ooyen, A., & Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. *Trends in Cognitive Sciences*, 14(2), 64–71.

<https://doi.org/10.1016/j.tics.2009.11.005>

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1.

<https://psycnet.apa.org/journals/rev/84/1/1/>

Schwabe, L., & Wolf, O. T. (2009). Stress Prompts Habit Behavior in Humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(22), 7191–7198.

<http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0979-09.2009>

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.

<http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=23889930&retmode=ref&cmd=prlinks>

Solley, C. M., & Murphy, G. (1960). Effects of Practice and Reward. In C. M. & M. Solley (Ed.), *Development of the perceptual world*, (pp (Vol. 353, pp. 81–103). Basic Books, xiv.

<https://doi.org/10.1037/11120-005>

Son, L. K., & Sethi, R. (2006). Metacognitive control and optimal learning. *Cognitive Science*, *30*(4), 759–774. [https://doi.org/10.1207/s15516709cog0000\\_74](https://doi.org/10.1207/s15516709cog0000_74)

Steinglass, J., & Walsh, B. T. (2006). Habit learning and anorexia nervosa: a cognitive neuroscience hypothesis. *The International Journal of Eating Disorders*, *39*(4), 267–275.

<https://doi.org/10.1002/eat.20244>

Sullivan-Toole, H., Richey, J. A., & Tricomi, E. (2017). Control and Effort Costs Influence the Motivational Consequences of Choice. *Frontiers in Psychology*, *8*, 675.

<https://doi.org/10.3389/fpsyg.2017.00675>

Swanson, S. D., & Tricomi, E. (2014). Goals and task difficulty expectations modulate striatal responses to feedback. *Cognitive, Affective & Behavioral Neuroscience*, *14*(2), 610–620.

<http://link.springer.com/article/10.3758/s13415-014-0269-8>

Tajima, S., Drugowitsch, J., & Pouget, A. (2016). Optimal policy for value-based decision-

- making. *Nature Communications*, 7(1), 1–12. <https://doi.org/10.1038/ncomms12400>
- Tversky, A., & Kahneman, D. (1986). Rational Choice and the Framing of Decisions. *The Journal of Business*, 59(4), S251–S278. <http://www.jstor.org/stable/2352759>
- Vandaele, Y., & Ahmed, S. H. (2021). Habit, choice, and addiction. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 46(4), 689–698. <https://doi.org/10.1038/s41386-020-00899-y>
- Versace, F., Robinson, J. D., Lam, C. Y., Minnix, J. A., Brown, V. L., Carter, B. L., Wetter, D. W., & Cinciripini, P. M. (2010). Cigarette cues capture smokers' attention: evidence from event-related potentials. *Psychophysiology*, 47(3), 435–441. <https://doi.org/10.1111/j.1469-8986.2009.00946.x>
- Voon, V., Baek, K., Enander, J., Worbe, Y., Morris, L. S., Harrison, N. A., Robbins, T. W., Rück, C., & Daw, N. (2015). Motivation and value influences in the relative balance of goal-directed and habitual behaviours in obsessive-compulsive disorder. *Translational Psychiatry*, 5(11), e670–e670. <https://doi.org/10.1038/tp.2015.165>
- Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, 411(6840), 953. <https://www-nature-com.stanford.idm.oclc.org/articles/35082081>
- Waskom, M. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. <https://doi.org/10.21105/joss.03021>
- Waskom, M. L., Kumaran, D., Gordon, A. M., Rissman, J., & Wagner, A. D. (2014). Frontoparietal Representations of Task Context Support the Flexible Control of Goal-Directed Cognition. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 34(32), 10743–10755. <https://doi.org/10.1523/JNEUROSCI.5282-13.2014>
- Waskom, M. L., Okazawa, G., & Kiani, R. (2019). Designing and Interpreting Psychophysical Investigations of Cognition. *Neuron*, 104(1), 100–112. <https://doi.org/10.1016/j.neuron.2019.09.016>

- Waskom, M. L., & Wagner, A. D. (2017). Distributed representation of context by intrinsic subnetworks in prefrontal cortex. *Proceedings of the National Academy of Sciences*, *114*(8), 2030–2035. <http://www.pnas.org/lookup/doi/10.1073/pnas.1615269114>
- Weintraub, D. (2008). Dopamine and impulse control disorders in Parkinson's disease. *Annals of Neurology*, *64 Suppl 2*(Suppl 2), S93–S100. <https://doi.org/10.1002/ana.21454>
- Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*, 14. <https://doi.org/10.3389/fninf.2013.00014>
- Wilson, R. C., & Collins, A. G. E. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, e49547. <https://doi.org/10.7554/eLife.49547>
- Wood, W. (2017). Habit in Personality and Social Psychology. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *21*(4), 389–403. <https://doi.org/10.1177/1088868317720362>
- Wood, W., & Runger, D. (2016). Psychology of Habit. *Annual Review of Psychology*, *67*(1), 289–314. <https://doi.org/10.1146/annurev-psych-122414-033417>
- Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews. Neuroscience*, *7*(6), 464–476. <http://www.nature.com/doi/10.1038/nrn1919>