Title

Reward Reinforcement Creates Enduring Facilitation of Goal-directed Behavior.

Permalink

https://escholarship.org/uc/item/5cn2m75f

Journal

Journal of Cognitive Neuroscience, 36(12)

Authors

Ballard, Ian

Waskom, Michael
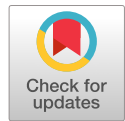
Nix, Kerry

et al.

Publication Date

2024-12-01

DOI

10.1162/jocn_a_02150

Peer reviewed

Reward Reinforcement Creates Enduring Facilitation of Goal-directed Behavior

Ian C. Ballard[1], Michael Waskom[2], Kerry C. Nix[3], and Mark D'Esposito[4]

Abstract

■ Stimulus–response habits benefit behavior by automatizing the selection of rewarding actions. However, this automaticity can come 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 initiation and execution of rules. Strikingly, we found that these biases persisted in a test phase in which rewards were no longer available. Although this facilitation is consistent with the habitual goal selection hypothesis, we did not find evidence that reward reinforcement reduced cognitive flexibility to implement alternative rules. Together, the findings suggest that reward creates a lasting impact on the selection and execution of goals but may not lead to the inflexibility characteristic of habits. Our findings demonstrate the role of the reward learning system in influencing how the goal-directed system selects and implements goals. ■

INTRODUCTION

Habits are powerful determinants of daily decisions and contribute to maladaptive behaviors in neurocognitive disorders (Wood & Rünger, 2016; Lhermitte, 1983). Habitual behavior is often characterized as a rote or automatic behavioral response to a specific stimulus, such as stopping at a red light (Knowlton, Mangels, & Squire, 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 (Voon et al., 2015; Gillan & Robbins, 2014), and in Parkinson 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 (Niv, 2009; Graybiel, 1998; DeLong, 1990). 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 pFC represent abstract rules and goals (Wallis, Anderson, & Miller, 2001) rather than actions but share a similar, overlapping corticostriatal architecture with motor cortex (Haber, 2011; Alexander, DeLong, & Strick, 1986). It has been hypothesized that reward reinforces abstract task representations analogously to cortical action representations (Radulescu, Niv, & Ballard, 2019; Collins & Frank, 2013; Badre & Frank, 2012; Frank & Badre, 2012; 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

[1]University of California, Riverside, [2]New York University, [3]University of Pennsylvania, [4]University of Califonia, Berkeley

task sets (Rmus, McDougle, & Collins, 2021; Eckstein & Collins, 2020; Collins & Frank, 2013). 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 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 that gave rise to the habit have changed. A key feature of goals is that they act in a context-dependent manner to influence behavior. For example, the way in which a person responds to their laptop will depend on the context of their goals. We operationalized the context-dependence of goal-directed behavior using a rule-based perceptual discrimination task. In this task, the way in which participants responded to a perceptual stimulus depended on the rule context. We found that reward reinforcement of rules influenced behavior in a manner consistent with the first and third predictions: Execution of the high-reward rule was improved, and this effect persisted after the opportunity to earn rewards was eliminated. However, we did not find evidence supporting Prediction 2: The ability to adapt behavior away from the high-reward rule was not reliably influenced by reward reinforcement. These results show that reward creates enduring impacts on the execution of rule-guided behavior but does not impair cognitive flexibility (de Wit et al., 2018).

METHODS

Participant 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 participants provided written informed consent, 65 are female participants, median age is 20 years, SD is 4.83 years, and range is 18–51 years. Data from the test blocks are missing from one participant because of a computer error. The target sample size, 85 participants, was chosen to have 80% power to detect a medium-sized correlation ($r = .3$) at an α of .05. Because we did not identify any outlier participants in behavioral performance (defined as 3 SDs below the mean accuracy) and all participants performed well-above chance, no participants were excluded.

Task Design

Participants performed a context-based perceptual discrimination task in which they could earn rewards for accurate performance (Waskom, Okazawa, & Kiani, 2019; Waskom & Wagner, 2017). On each trial, participants responded based on one of three rules, color, shape, or motion direction of a field of colored, moving shapes.

These rules capture the context-dependence of goal-directed behavior by determining which dimension of the stimulus must be attended to generate an appropriate response. 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. Participants were given up to 2 sec 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 2 sec regardless of when the participant responded. All three rules shared the same keys, that is, 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, $F(1.73, 147) = 20$, $\eta^2 g = .059$, $p < .001$, 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. Participants were instructed that, during the reward phase, some trials carried the potential to earn rewards for correct responses. Incorrect responses prevented the participant from earning a reward. One rule was randomly chosen for each participant 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 participants 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. We chose to manipulate reward probability, rather than reward magnitude, because of theoretical and empirical work suggesting that probabilistic rewards optimally drive learning (Wilson, Shenhav, Straccia, & Cohen, 2018). Participants performed six blocks of 96 trials for 572 trials of the reward task. Participants were told that one of the reward blocks would be selected randomly to count for real, and rewards from that block, each worth $0.50, would be paid as a bonus. We chose to make rewards contingent on accuracy,

rather than speed and accuracy, because, due to the difficulty of the task, incentivizing fast responses would have resulted in lower reward rates and could have reduced the salience of our reward manipulation.

Rewards were signaled by the color of the fixation cross changing to gold for 500 msec. On correct, unrewarded trials, there was no feedback. The fixation cross turned red on error trials, but participants were not penalized for incorrect responses. The task is very difficult on low coherence trials, and we did not want to penalize participants for making incorrect responses that are expected in the task. Otherwise, the task may have been much less rewarding overall. However, we wanted participants to stay on task, so we deducted rewards for missed responses. The fixation cross flickered red on trials where the participant failed to respond during the stimulus window, indicating to participants that a reward was deducted. Feedback was presented 300 msec after the offset of the stimulus. The intertrial interval after the feedback offset, or kinematogram offset on trials with no feedback, was 1000 msec.

Immediately after the reward phase, participants took an enforced 6-min break before beginning the test phase. The participants were instructed verbally at the beginning of the session that the test block was identical to the reward blocks, but there was no opportunity for rewards. We intended for the break to create a clear boundary between the phases and to provide an opportunity to consolidate reward learning (Murty, Alexa Tompary, Adcock, & Davachi, 2017). During the break, the task instructions on the screen said: "These blocks will not have rewards, but please try your hardest." Participants performed two blocks of 96 trials of extinction for 192 trials.

To measure the effect of rule habit on cognitive flexibility, rule order was presented in an unsignaled miniblock structure. Within each miniblock, participants 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. Participants were not instructed on the miniblock structure.

Participants were trained on the behavioral task in a 2-hr session 1–3 days before the main task. Participants 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 three-down-one-up staircase (i.e., the coherence was reduced after three consecutive correct responses and increased after every error).

Subsequently, participants 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, participants 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., RT), mixed-effects models were implemented using the *lmer* package in R 4.2.0 (Bates, Mächler, Bolker, & Walker, 2015). Because RTs are positive and non-Gaussian, RTs were log-transformed before being entered as dependent variables. 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 participant 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, Levy, Scheepers, & Tily, 2013). Mixed-effects models with random slopes for rule type failed to converge and were therefore removed in our model. Data plots were created using Seaborn 0.11.2 (Waskom, 2021).

Drift-diffusion Modeling

The underlying decision-making processes in our task were assessed using a drift-diffusion model (DDM). The DDM characterizes decision-making as noisy evidence accumulation over time. In our DDM models, there is a variable period before evidence initiation begins (*initiation time*), after which evidence noisily accumulates toward the correct response with an average rate determined by the drift rate. Once evidence for a response passes a threshold, determined by the decision boundary, a response is made. Drift-diffusion modeling was performed using HDDM 9.2 (hierarchical drift-diffusion model) (Wiecki, Sofer, & Frank, 2013). This hierarchical Bayesian model allows simultaneous estimation of model parameters for the entire group of participants while constraining parameter estimates for individual participants. Models were fit independently for the reward reinforcement and extinction test phases using Markov-chain Monte Carlo with 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 r-hat < 1.01) for all parameters, indicating that the five chains

converged to the same stationary distribution. Our models assumed that each participant'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 participants to derive predicted accuracy and RTs.

To test different hypotheses about the impact of reward in our task, we fit six different models that varied in which parameters of the drift-diffusion process were affected by rule type. We modeled the effects of experimental conditions on drift-rate, decision threshold, and initiation time parameters using a within-subject regression model, which allowed us to account for individual differences in overall task performance by estimating individual participant intercepts for each DDM parameter. Our winning model, which includes the effect of experimental condition on drift rate, decision threshold, and initiation time, is specified as:

Drift rate: $1 + \theta_c$ *coherence* $+ \theta_{drift\_rule}$rule_type.
Decision Threshold: $1 + \theta_{threshold\_rule}$rule_type.
Initiation time: $1 + \theta_{initiation\_rule}$rule_type.

Rule type (*rewarded, competing*, and *noncompeting*) was modeled using contrast coding with *noncompeting* rules set as the baseline. The slope parameters of the regression model, θ, were fixed effects whereas the intercepts were random effects. The regression model parameters were fit jointly with the default parameters of the hierarchical DDM model. The other five models we tested varied in whether rule type influenced drift rate, decision threshold, or threshold (1: drift rate only, 2: decision threshold only, 3: initiation time only, 4: drift rate and decision threshold, 5: drift rate and initiation time). In all models, the coherence of the kinematogram influenced the drift-rate parameter (θ_c).

RESULTS

The participants performed a rewarded, context-based perceptual discrimination task (see 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 participants were not instructed on reward contingencies, they were more accurate, $z = 6.9, p < .001$, odds ratio = 1.19 (Figure 2A), and faster, $t(48840) = 2.01, p = .044, \Delta RT = .41\%$ (Figure 2B), on high-reward relative to low-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 participants were instructed that there was no possibility of reward. We predicted that 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, $z = 3.6, p < .001$, odds ratio = 1.15 (Figure 2A), and RTs were faster, $t(15800) = 4.12, p < .001, \Delta RT = 1.73\%$ (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.

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 (Roelfsema, van Ooyen, & Watanabe, 2010; Law & Gold, 2008; Solley & Murphy, 1960). We sought to minimize the influence of perceptual learning by training participants on the perceptual discrimination task on a previous training day. We asked whether there was evidence of continued perceptual learning during the reward phase in spite of this training. We modeled potentially nonlinear learning effects by examining the linear and quadratic effect of trial number on accuracy, as well as the interaction between these linear and quadratic trial number effects with rule type. We found trending evidence for both a linear, $Z = 1.82, p = .068$, odds ratio = 1.03, and quadratic, $Z = -1.85, p = .064$, odds ratio = 1.03, effect of trial number on accuracy. However, perceptual learning was not different for the high-reward rule versus low-reward rule during learning, $ps > .1$ for linear and quadratic interactions. Therefore, although perceptual learning may have continued during the task, it did not differ between rule types and therefore does not account for 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 (Waskom et al., 2019). As expected, coherence strongly affected rule accuracy, $z = 24, p < .001$, odds ratio = 1.74, and RT, $t(48840) = -19.5, p < .001, \Delta RT$ per unit coherence = 4.56%, 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 $Z = -4.3, p < .001$, odds ratio = 1.10; RT, $t(48840) = 6.0, p < .001, \Delta RT = 1.18\%$, it was opposite to the predicted direction: We found that reward had the largest impact on easier trials.
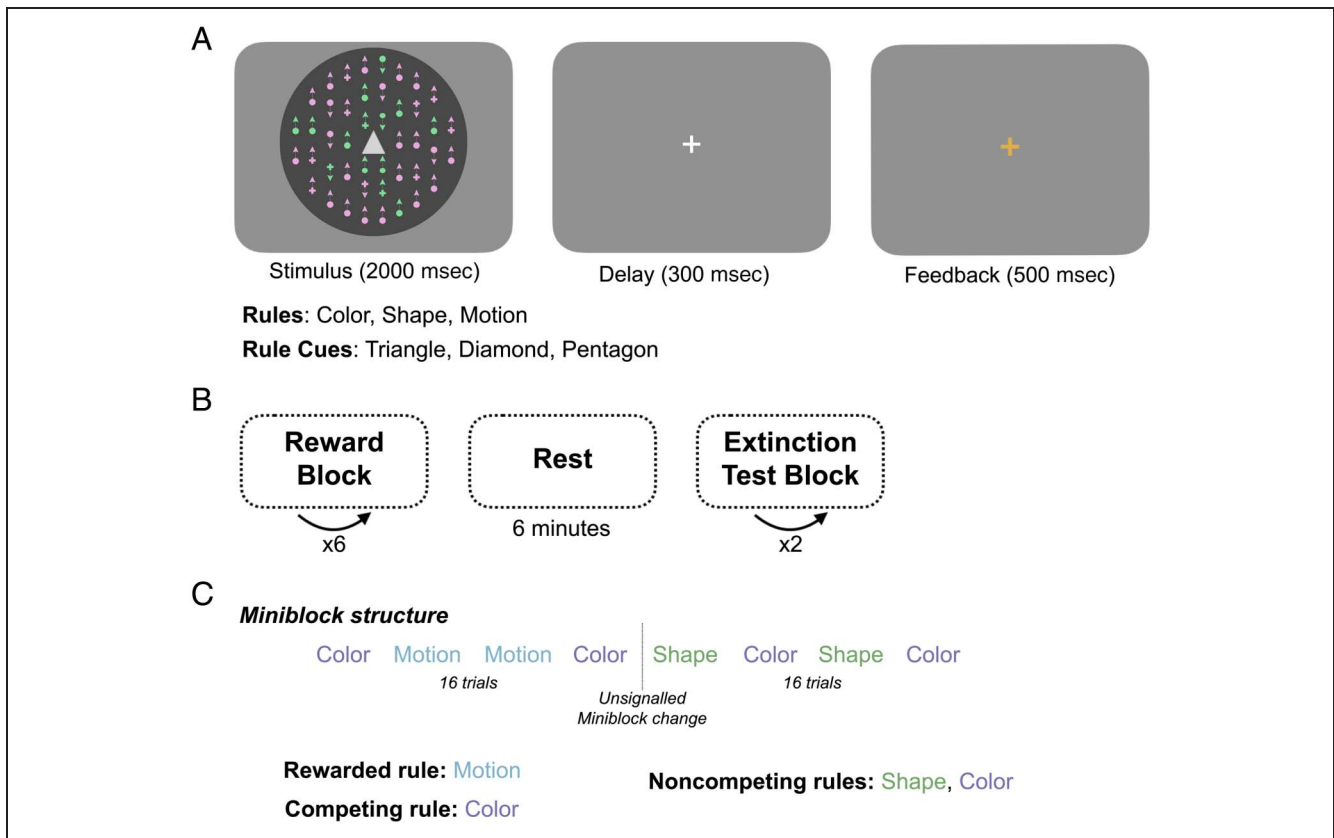
Figure 1. Task design. (A) Participants ($n = 86$) performed a rewarded, context-dependent, perceptual decision-making task. On each trial, a central cue (a triangle in the above example) indicated whether participants responded based on the shape, color, or motion of a shape kinematogram. Accurate responses on one of the rules were rewarded at a higher rate (85%) than the other two rules (15%). Feedback indicated whether the participant 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, participants 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 noncompeting rule in a (shape, color) miniblock.

One account of this finding is that participants stood to gain the most reward with the least cognitive effort by improving performance on easy trials (Shenhav, Botvinick, & Cohen, 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, $z = 2.46$, $p = .014$, odds ratio $= 1.09$, and a trending interaction for RTs, $t(15800) = 1.65$, $p = .092$, $\Delta RT = 0.63\%$. In the following section, we use a DDM approach to provide an alternative account of this behavioral effect.

Habits improve the execution of rewarding behaviors at the cost of reduced flexibility to adapt behavior when goals change. Importantly, this reduction in flexibility should only occur when habits compete against alternative behaviors for control of behavior. 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, participants 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. We tested whether competition with a rewarded rule impaired rule execution. During the reward phase, accuracy varied as a function of Rule Coherence, $F(2.5, 210) = 709$, $\eta^2 g = .62$, $p < .001$; Rule Type (high-reward, competing, noncompeting), $F(1.4, 121) = 6.4$, $\eta^2 g = .016$, $p = .006$; and a trend toward an interaction between Rule Type (high-reward, competing, noncompeting) and Coherence, $F(5.3, 452) = 2.1$, $\eta^2 g = .006$, $p = .061$ (Figure 3A). These relationships persisted into the test phase, where we found a main effect of Rule Coherence, $F(2.5, 210) = 215$, $\eta^2 g = .34$, $p < .001$; Rule Type (high-reward, competing, noncompeting), $F(1.9, 157) = 3.6$, $\eta^2 g = .006$, $p = .033$; and an interaction between Rule Type and Coherence, $F(5.3, 448) = 2.8$, $\eta^2 g = .011$, $p = .015$. However, contrary to our
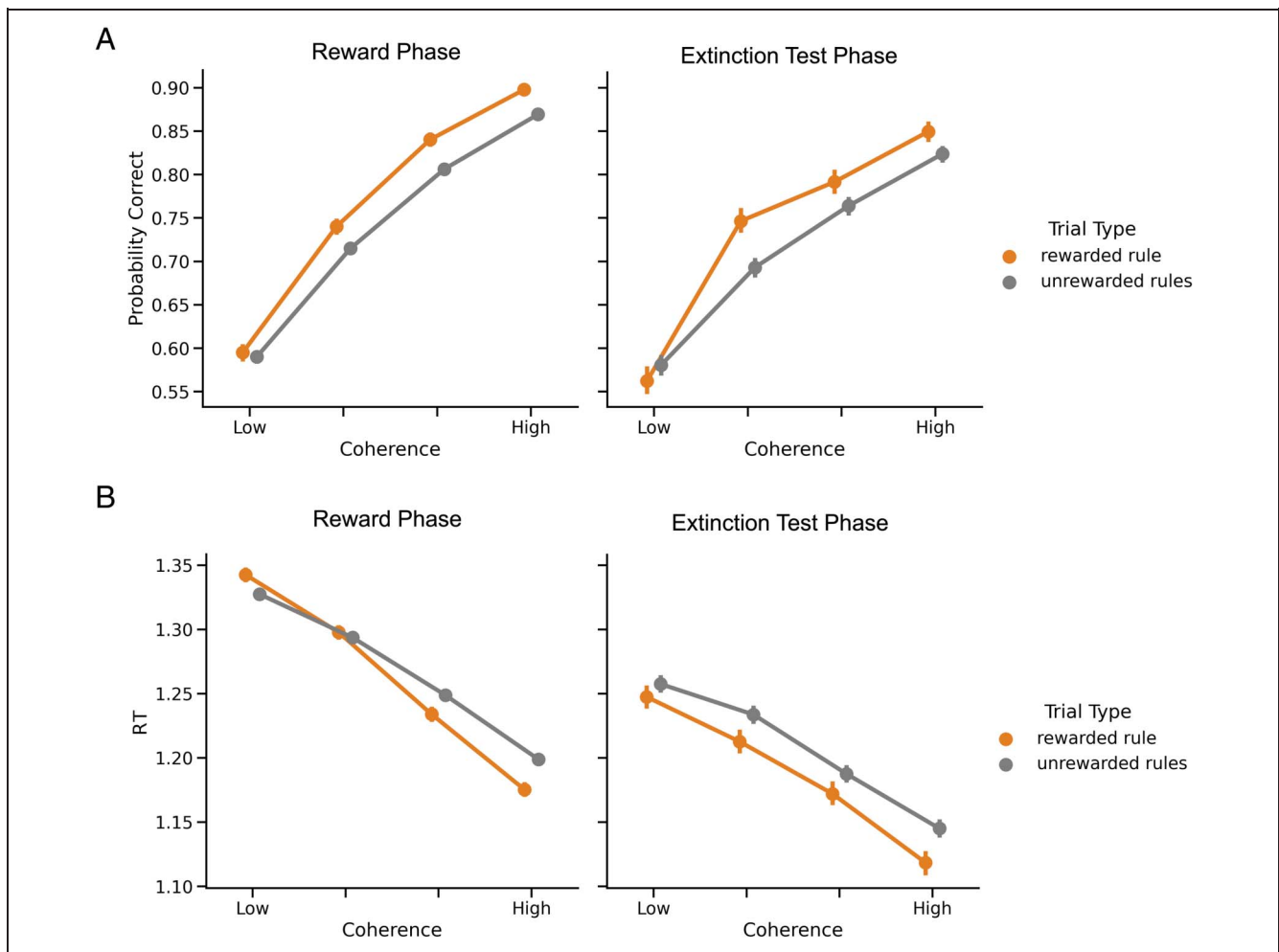
Figure 2. Reward reinforcement creates rule habits. (A) Accuracy was higher for the high-reward rule than the low-reward rules in both the reward and test phases. In addition, accuracy was higher for high-coherence (easy) trials than for low-coherence (hard) trials. (B) RTs were faster for the previously high-reward rule in the test phase. During the reward phase, there was an interaction between coherence and rule type on accuracy and RTs.

predictions, there was no difference in accuracy between competing and noncompeting rules in the reward nor the test phases, $ps > .2$. Consistent with impaired flexibility, we found that RTs were faster for the noncompeting, relative to the competing rule during the reward phase, $t(48840) = 6.17, p < .001, \Delta RT = 1.45\%$ (Figure 3C). However, this effect was not present in 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 RT 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. In addition, 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, noncompeting) as regressors. The trial-type regressor ensures that any differential switch costs are not simply because of performance differences between the rules. We found that during the reward phase, participants were faster at switching to the high-reward rule, relative to switching between low-reward rules, $t(48840) = -5.4, p < .001, \Delta RT = 2.45\%$. However, in contrast to the predictions of the rule habit hypothesis, 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, whereas the flexibility to adapt behavior is unaffected by reward. Similarly to the previous results, the switch cost analysis showed that reward reinforcement improves rule selection without affecting cognitive flexibility.
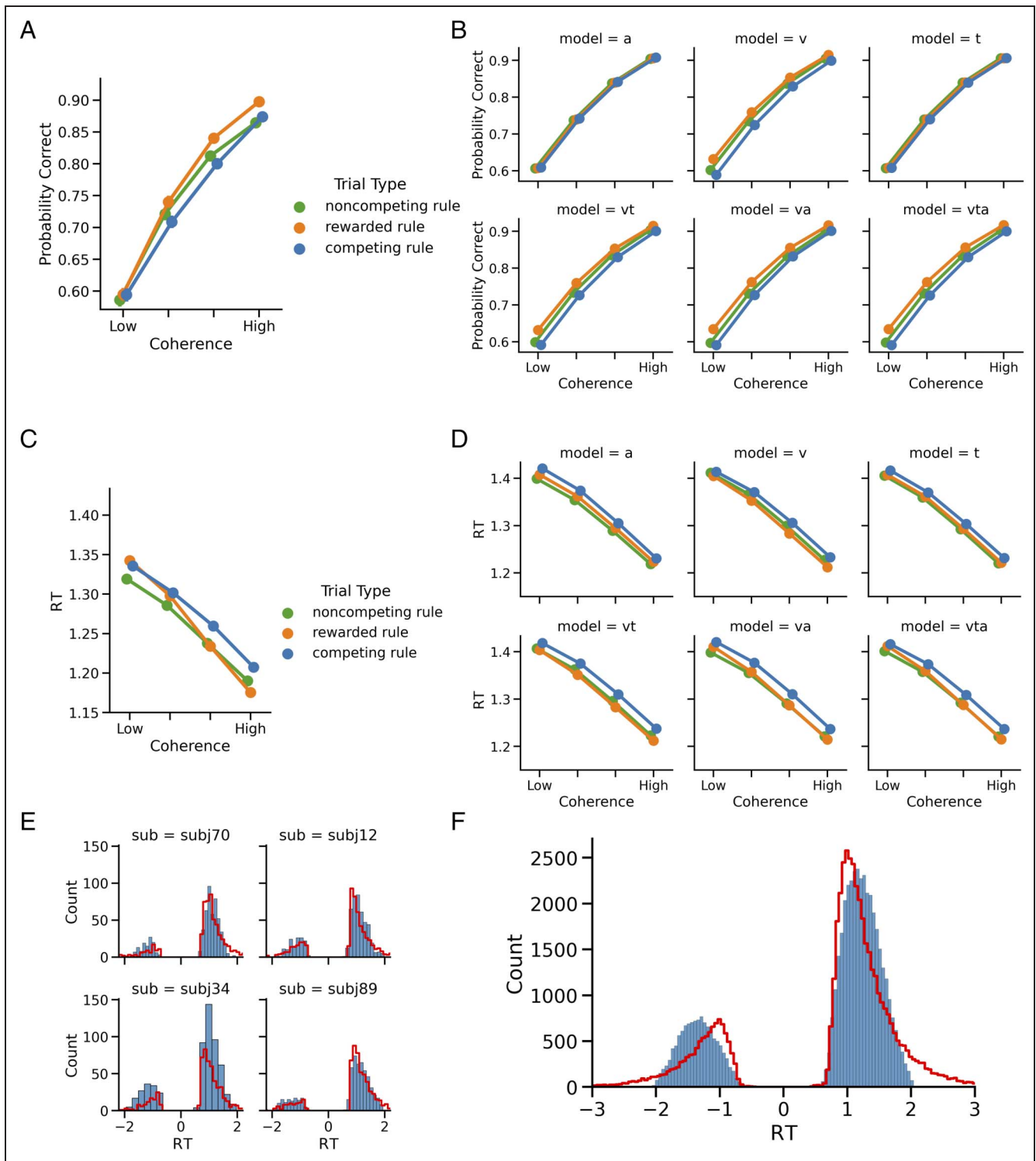
Figure 3. Qualitative model comparison. Models are labeled by parameters that were influenced by rule condition: *a = decision threshold, t = initiation time*, and *v = drift rate.* (A) Participant accuracy by rule coherence and experimental condition for the reward phase. Participants are more accurate 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) Participant RTs by rule coherence and experimental condition for the reward phase. Participants are faster for the noncompeting relative to the competing rule. In addition, RTs for the rewarded rule are faster on easy relative to hard trials. (D) Simulated RT data for each model. Most models could capture the RT difference between competing and noncompeting 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 RT. (E) Model fits for four randomly chosen participants. Empirical RT distributions are shown in blue; simulated RT distributions are traced in red. (F) Model fits for the entire cohort. Empirical RT distributions are shown in blue; simulated RT distributions are traced in red. The model predicts RTs that are slower than our response threshold of 2 sec.

Drift Diffusion Models of Choice

Reward Influences Multiple Components of Rule-guided Behavior

We found evidence that reward reinforcement influenced the speed and accuracy of rule execution even after rewards were eliminated. This influence of reward on rule performance could arise from multiple different mechanisms. 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, Meijer, & Pennartz, 2018; Hickey, Kaiser, & Peelen, 2015) or by shifting the speed-accuracy tradeoff in favor of accuracy (Tajima, Drugowitsch, & Pouget, 2016; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006). We fit a DDM to our data to discover how reward influences rule-guided behavior. 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 RTs. 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 RTs. This parameter can capture variability in response caution between rule types.
3) The *nondecision time* captures the time needed to initiate the drift-diffusion process. Conditions with higher nondecision time will have slower RTs without necessarily having higher accuracy. This parameter can capture variability in the time it takes to select a rule.

We first sought to establish whether rule type influenced each of these parameters. Our model selection strategy employed a balanced consideration of both the goodness-of-fit, the *deviance information criterion* (DIC), and posterior 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 (Figure 3A); (2) the interaction between coherence and rule type on accuracy (Figure 3B) and RT (Figure 3C);

and (3) the slower RTs for the competing, relative to the noncompeting rule (Figure 3C).

We first examined three models in which rule type (high-reward, competing, or noncompeting) influenced only a single parameter: either drift rate, decision threshold, or nondecision time. Of these models, only the drift-rate model could accurately capture the higher accuracy for the high-reward rule trials (Figure 3B). However, this model failed to capture the other two behavioral effects (Figure 3D), indicating that a more complex model was needed to explain our behavioral data.

We next asked whether including additional effects of rule type on decision threshold or nondecision time to the drift-rate model could capture all three behavioral effects of interest. We found that both the (drift-rate, threshold) models and the full (drift-rate, threshold, nondecision time) models could capture all three qualitative behavioral effects (Figure 3B and D). We decided to use the (drift-rate, threshold, nondecision time) model as the final model for our data for three reasons:

1. The posterior predictive checks show a small but significant improvement in the model's ability to account for the data (Figure 3D).
2. It had the lowest DIC (Figure 4D), 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 nondecision time).

To test the robustness of the model, we performed a parameter recovery study. We simulated a data set from the HDDM model using the average posteriors of the group-level parameters and fit a new model to that simulated data set. The recovered and ground truth parameters were highly correlated, $r = .99$, $p < .001$, and model fitting introduced an average squared error of only .051% of the ground-truth parameters. These positive model recovery results conform with other results showing that the HDDM toolbox gives meaningful parameter estimates (Wiecki et al., 2013).

Our final model explains key qualitative features of our data; however, the model overestimates the mean accuracy and RTs in our data (Figure 3D, E, and F). The overestimation of RTs occurs because participants were forced to respond within 2000 msec, whereas the DDM is a continuous model that predicts some RTs slower than this threshold. In addition, there was an underestimation of RTs on error trials. This arises because drift diffusion models predict symmetric correct and error RT distributions, whereas humans tend to respond more slowly on error trials in difficult tasks (Ratcliff & Rouder, 1998). HDDM models with trial-by-trial variability in parameters estimates can theoretically account for this pattern; however, we were unable to reliably fit these highly complex
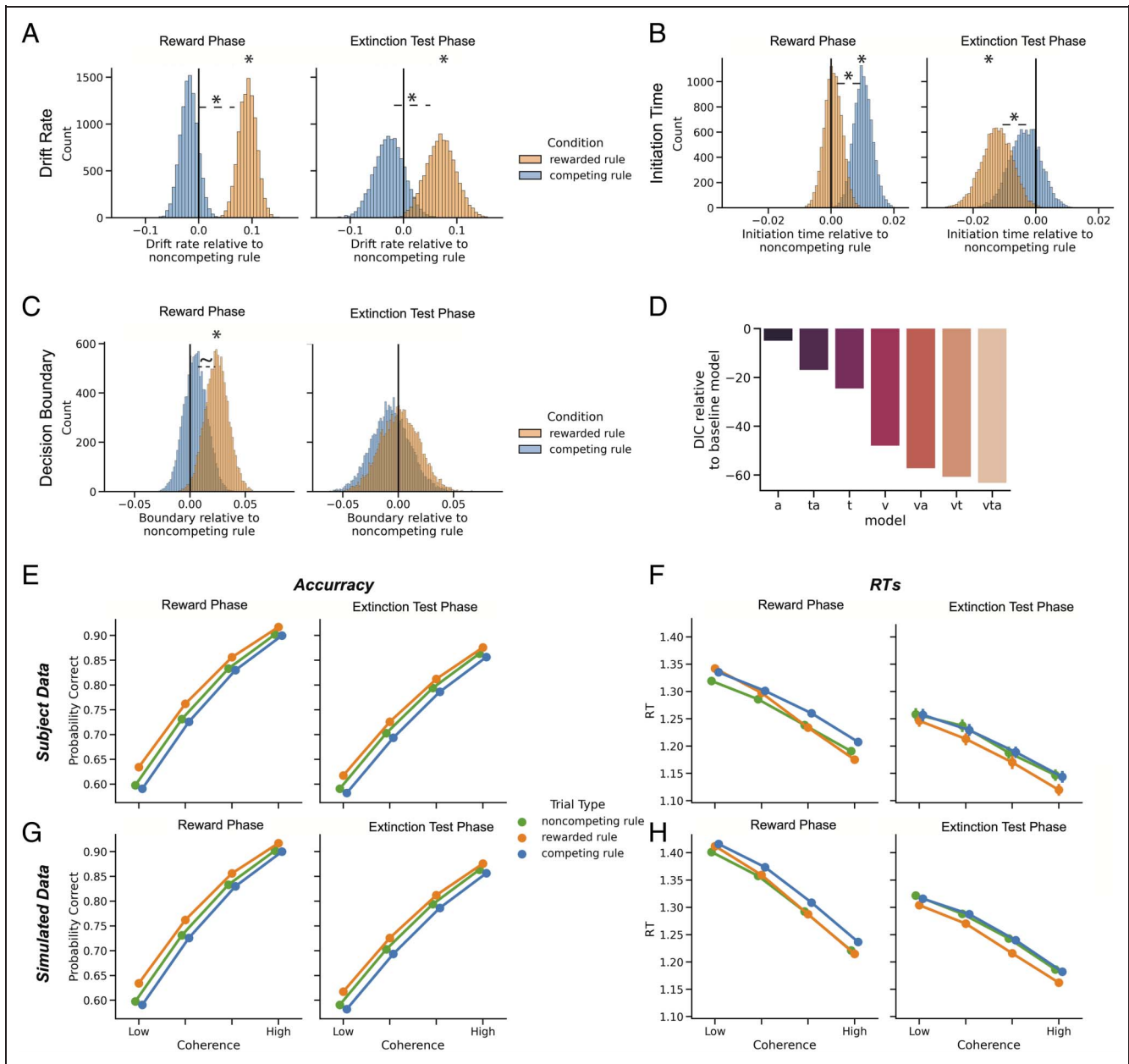
Figure 4. Reward reinforcement facilitates rule initiation and execution. (A) Drift rates were higher for the high-reward rule during both reward reinforcement and the test phase, indicating persistent enhancement of rule implementation by reward. (B) Initiation times, which reflect the time it takes to begin evidence integration, were slower for the competing rule during the reward phase only. Initiation times were faster for the high-reward rule relative to other rules in the test phase only. (C) Decision thresholds, which control the speed-accuracy tradeoff, were increased for the high-reward rule during the reward-phase only. (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) Participant accuracy and (F) RTs for each rule condition are well matched by simulated data (G–H) from the model. * Indicates evidence or strong evidence; ~ indicates trending evidence. Error bars depict standard error (E–G).

models to our data. Critically, the observed distributions of participants' accuracy and RTs are within the credible intervals (CIs) predicted by the model, indicating that the model considers the empirical data likely under its parameterization. Moreover, our model accounts for, and can reproduce in simulation, the prominent behavioral effects in our data described above. Therefore, although the model did not fully account for the shape of the empirical RT distribution, the quantitative and qualitative evaluation of the model described above establish that the model explains important features of our data and has stable parameter estimates (Wilson & Collins, 2019). We

next examined which latent processes in the model accounted for our observed behavioral effects.

Impacts of Reward on Rule Execution

We hypothesized 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 probabilities of direction *(pd)* indicate the model's evidence that an effect exists, given the data. We considered *strong evidence* to be a *pd* of 99.9% or higher, *evidence* to be *pd* was greater than 99%, and *trending evidence* to be a *pd* of 95% or higher. We additionally report 89% CIs for parameter estimates (Kruschke, 2014). We found strong evidence that the drift rates were higher for the high-reward rule relative to the noncompeting rule, *pd* = 100%, CI [.065, .12], and the competing rule, *pd* = 100%, CI [.08, .13]. If improved execution of the rewarded rule was entirely driven by reward motivation, we would not expect this facilitation to persist. However, we found that facilitation of the high-reward rule persisted into the test phase, *pd* of reward > noncompete = 99.6%, CI [.027, .12], reward > compete = 100%, CI [.051, .14] (Figure 4A). Moreover, there was no statistical difference between drift rates between the reward and test phases. These results show that reward creates an enduring facilitation of rule execution.

Because habits can impair the flexibility to change behavior, we next examined whether execution of rules that compete against high-reward rules was impaired. However, we again did not find evidence that drift rates were lower for competing rules than noncompeting rules in the reward phase, *pd* = 86.9%, nor the test phase, *pd* = 83.3% (Figure 4A). Similarly to the behavioral effects, we did not find evidence that reward reinforcement influences the 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 participants 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, the impact of response incongruency (when the dimensions of the kinematogram

indicated conflicting button responses) was reduced for the high-reward rule, relative to the other rules, *t*(48890) = −2.0, *p* = .049, ΔRT = .94%. 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 before rule execution in this task. We found strong evidence that the initiation time was reduced for the reward rule relative to the noncompeting rule during the test phase, *pd* = 99.6%, *pd* = [−.02, −.005], but not the reward phase. A direct comparison of the task phases revealed that initiation times were faster for the rewarded rule in the test phase than the reward phase, *pd* = 99.1%, *pd* = [.004, .023]. 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 participants were slower to initiate low-reward rules that were competing in a context with high-reward rules. During the reward phase, we found strong evidence that the initiation time of the competing rule was increased relative to the noncompeting rule, *pd* = 100%, *pd* = [.005, .014] (Figure 4B). 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 noncompeting rule, *pd* = 75.2% (Figure 4B) and were reduced relative to the reward phase, *pd* = 99.4%, *pd* = [.004, .022]. These findings suggest 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 have a priori predictions about the influence of habit on response caution. However, during the reward phase, participants 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 RT (Tajima et al., 2016; Bogacz et al., 2006). Consistent with this prediction, we found evidence that the decision thresholds were higher for the high-reward rule relative to the noncompeting rule, *pd* = 98.8%, CI [.006, .04], and trending evidence relative to the competing rule, *pd* = 95.7%, CI [.001, .036] (Figure 4C). Because this

adjustment of the decision threshold is adaptive for earning rewards, we did not expect it to persist into the test phase. Although rule condition did not influence decision thresholds in the test phase, *pds* < 70%, decision thresholds did not significantly differ between reward and test phases for the rewarded nor the competing rule (Figure 4C). In addition, there were no differences between the competing and noncompeting rules in either task phase, *pds* < 70%. In summary, participants adjusted their response caution adaptively, responding more carefully only on trials where rewards were likely (Grahek, Schettino, Koster, & Andersen, 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 RTs only on easier trials (Figure 4F, left). The DDM shows that this effect arises because reward influences both drift rates and decision thresholds. Increased drift rates lead to overall faster RTs for high-reward rules. However, increased decision thresholds cause slower RTs for harder, high-reward rules. Together, these factors predict that the reward will cause the fastest RTs on easy trials, the effect found in our data (Figure 4H). 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, Shenhav, & Botvinick, 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, Jessup, & Tricomi, 2017; Sullivan-Toole, Richey, & Tricomi, 2017).

DISCUSSION

We found evidence that reward reinforcement leads to enduring facilitation of the selection and execution of rules. 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 participants 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 (Shenhav et al., 2013; Ballard et al., 2011; Niv, Daw, Joel, & Dayan, 2007). Another potential mechanism is that reward strengthens the associative link between the cue and its associated rule representation (Miller, Freedman, & Wallis, 2002). This strengthened association could facilitate activation of the rule representation, which would account for the finding that participants 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 rule-relevant dimension (Etzel, Cole, Zacks, Kay, & Braver, 2016; Waskom, Kumaran, Gordon, Rissman, & Wagner, 2014). This interpretation is supported by the finding that response incongruency across the dimensions of the stimulus (e.g., color and motion dimensions indicating opposite button responses) had a smaller effect on RTs for the rewarded rule. We posit that reward influences the deployment of top–down attention to facilitate sensory evidence integration (Frömer, Lin, Dean Wolf, Inzlicht, & Shenhav, 2021; Krebs & Woldorff, 2017; Botvinick & Braver, 2015). 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 (Failing & Theeuwes, 2014; Anderson, Laurent, & Yantis, 2011; Della Libera & Chelazzi, 2009). 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 (Wood & Rünger, 2016; Versace et al., 2010).

Our drift-diffusion modeling approach revealed that reward reinforcement differentially influenced different components of rule-guided behavior. Reward had a strong influence on drift rates, which captures rule execution, during both reward learning and the test phase. In addition, reward reduced nondecision times, which may capture rule selection time, in the test phase. Although the nondecision time effect was not found in the reward phase, nondecision times for the rewarded rule did not differ meaningfully between the phases. In contrast, reward influenced decision threshold, which reflects response caution, during the reward, but not the test phase. We interpret the decision threshold effect as a strategic adjustment of the speed-accuracy tradeoff to earn rewards in a

task that incentivizes accuracy. Unlike the influence of reward on drift rate, which also improves task performance, this decision threshold effect was reduced when rewards were eliminated. This suggests that the mechanisms by which reward influences these processes differ: Whereas the adjustment of response caution may be strategic and context-dependent (Frank, 2006; Son & Sethi, 2006), the enhancement of rule execution may arise because of reward reinforcement of prefrontal rule representations (O'Reilly & Frank, 2006). A prediction arising from this interpretation is that in designs where speed, rather than accuracy, is incentivized, then reward will reduce response caution while also improving rule execution.

One potential weakness of our modeling results is that the DDM overestimated mean accuracy and RTs, primarily because of the fact that the HDDM predicted RTs slower than our response threshold. Importantly, the empirical accuracy and RTs were within the CIs of the model, indicating that the model considered the observed data likely. In addition, the model was robust in a parameter recovery study and explained several important qualitative features of our data (Wilson & Collins, 2019; Box, 1976). Nonetheless, it will be important to develop adapted DDMs in future work that can incorporate response thresholds and more accurately capture the shape of the RT distributions in our data.

We did not find consistent evidence that reward reinforcement reduced the flexibility to adapt behavior. Although we observed that RTs were reduced for the noncompeting relative to the competing rule during the reward phase of the experiment, this effect did not persist into the test phase. One possible explanation is that reward does reduce cognitive flexibility, but this effect does not persist once rewards are removed. However, 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 noncompeting miniblocks, the lower reward rate could render the small reward probability of the noncompeting rules relatively more valuable. This account is consistent with our finding that there was no difference in initiation time between the competing and noncompeting rules in the test phase, when the reward rates of the two conditions were matched. Future work will be needed to disentangle the range adaptation account from the cognitive flexibility account of these competition effects.

Our study may have failed to find reduced flexibility associated with reward reinforcement for several reasons. First, it is important to note that the competition analyses compare performance on the same rule in different contexts. It is possible that our study was not sufficiently powerful to detect small effects of competition on rule execution. However, our study is also consistent with the dearth of evidence for experimental induction of inflexible stimulus–response habits. De Wit and colleagues presented five studies that fail to show that overtraining instrumental behavior results in inflexible motor habits (de Wit et al., 2018). These studies used variants of outcome devaluation paradigms, in which the outcome associated with a previously valuable stimulus is rendered less valuable and behavior is tested. Although rodents that are overtrained in outcome devaluation paradigms persist in selecting the stimulus associated with the devalued outcome (Tricomi, Balleine, & O'Doherty, 2009; Dickinson, Balleine, Watt, Gonzalez, & Boakes, 1995), humans failed to show such effects across five studies (de Wit et al. 2018, but see Tricomi et al., 2009).

Why are inflexible habits so difficult to detect in laboratory settings? One possible explanation is that much more extensive training across months is needed to establish a habit (Lally, van Jaarsveld, Potts, & Wardle, 2010). Another possibility is that habits trigger preparation of responses, but the cognitive control abilities of humans enable effective suppression of habitual behavior in laboratory conditions. Hardwick and colleagues asked participants to learn a visuomotor association over 4 days of training and then asked whether that association would interfere with implementation of alternative behaviors. They found that expression of the habit occurred only when participants were forced to respond rapidly (300–600 sec), presumably before cognitive control could inhibit the habitual response (Hardwick, Forrence, Krakauer, & Haith, 2019). Similarly, Sternberg and colleagues found that blocking, a prominent reward conditioning effect observed in rodents, only occurs in humans when they are forced to respond rapidly (Sternberg & McClelland, 2012). Given these findings in the stimulus–response domain, it will be important for future work to test the theory that reward reinforcement can cause inflexible cognitive habits by employing paradigms that either extend training over much longer timescales or involve rapid probes that can assess prepotent goal selection.

Our measures showed persistent facilitation of the rewarded rule in a test period that occurred several minutes after reward learning. However, it is unclear the timescale over which this facilitation lasts. Our 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). Future work is needed to explore the interacting psychological conditions, including temporally extended learning and reward anticipation (Ballard, Hennigan, & McClure, 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 effects of reward on goal-directed behavior.

Our results show that reward reinforcement creates persistent facilitation of rule-guided behavior. However, significant differences exist between the constructs in our task and real-world goals. Most significantly, whereas goals can involve flexible pursuit of different strategies (e.g., an exercise goal may involve swimming, running, and weight training), in our rule task, the rules that participants could follow are fixed. Reward likely influences both the selection of goals and the strategies that people employ to pursue them, and future research is needed to examine how rewards influence strategy selection. 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, Ballard, Baribault, Bishop, & Collins, 2021; Swanson & Tricomi, 2014) in forming cognitive habits. The brain's reward system is involved in a variety of disorders, including anxiety (Lago, Davis, Grillon, & Ernst, 2017; Packard, 2009), obsessive-compulsive disorder (Gillan et al., 2014; Gillan & Robbins, 2014), anorexia nervosa (Foerde et al., 2021; Steinglass & Walsh, 2006), and Parkinson disease (Cools, 2011; Dubois & Pillon, 1997). Understanding the role of reward learning in these disorders will require an account of whether and why the dopaminergic system reinforces maladaptive goals.

Acknowledgments

Corresponding author: Ian C. Ballard, Psychology Department, University of California, 900 University Ave., Riverside, CA 92521, or via e-mail: iancballard@gmail.com.

Data Availability Statement

Analysis code and behavioral data are available at: https://github.com/iancballard/JOCN_Ballard_et_al_2024.

Author Contributions

Ian C. Ballard: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Visualization; Writing—Original draft; Writing—Review & editing. Michael Wascom: Conceptualization; Methodology; Writing—Review & editing. Kerry C. Nix: Formal analysis; Investigation; Methodology; Software; Writing—Review & editing. Mark D'Esposito: Conceptualization; Funding acquisition; Methodology; Supervision; Writing—Review & editing.

Funding Information

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this paper report its proportions of citations by gender category to be: M/M = .537; W/M = .204; M/W = .093; W/W = .167.

REFERENCES

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, PubMed: 3085570

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences, U.S.A.*, *108*, 10367–10371. https://doi.org/10.1073/pnas.1104047108, PubMed: 21646524

Badre, D., & Frank, M. J. (2012). Mechanisms of hierarchical reinforcement learning in cortico-striatal circuits 2: Evidence from fMRI. *Cerebral Cortex*, *22*, 527–536. https://doi.org/10.1093/cercor/bhr117, PubMed: 21693491

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*, *29*, 793–804. https://doi.org/10.1162/jocn_a_01098, PubMed: 28129051

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. *Journal of Neuroscience*, *31*, 10340–10346. https://doi.org/10.1523/JNEUROSCI.0895-11.2011, PubMed: 21753011

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*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001, PubMed: 24403724

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*, 700–765.

https://doi.org/10.1037/0033-295X.113.4.700, PubMed: 17014301

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, PubMed: 25251491

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*, 791–799. https://doi.org/10.1080/01621459.1976.10480949

Brewer, J. A., & Roy, A. (2021). Can approaching anxiety like a habit lead to novel treatments? *American Journal of Lifestyle Medicine*, *15*, 489–494. https://doi.org/10.1177/15598276211008144, PubMed: 34646095

Chiew, K. S., & Braver, T. S. (2014). Dissociable influences of reward motivation and positive emotion on cognitive control. *Cognitive, Affective, & Behavioral Neuroscience*, *14*, 509–529. https://doi.org/10.3758/s13415-014-0280-0, PubMed: 24733296

Collins, A. G. E., & Frank, M. J. (2013). Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychological Review*, *120*, 190–229. https://doi.org/10.1037/a0030852, PubMed: 23356780

Cools, R. (2011). Dopaminergic control of the striatum for high-level cognition. *Current Opinion in Neurobiology*, *21*, 402–407. https://doi.org/10.1016/j.conb.2011.04.002, PubMed: 21531543

Cools, R., & D'Esposito, M. (2011). Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biological Psychiatry*, *69*, e113–e125. https://doi.org/10.1016/j.biopsych.2011.03.028, PubMed: 21531388

Cushman, F., & Morris, A. (2015). Habitual control of goal selection in humans. *Proceedings of the National Academy of Sciences, U.S.A.*, *112*, 13817–13822. https://doi.org/10.1073/pnas.1506367112, PubMed: 26460050

DeLong, M. R. (1990). Primate models of movement disorders of basal ganglia origin. *Trends in Neurosciences*, *13*, 281–285. https://doi.org/10.1016/0166-2236(90)90110-V, PubMed: 1695404

de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., et al. (2018). Shifting the balance between goals and habits: Five failures in experimental habit induction. *Journal of Experimental Psychology: General*, *147*, 1043–1065. https://doi.org/10.1037/xge0000402, PubMed: 29975092

Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., & Boakes, R. A. (1995). Motivational control after extended instrumental training. *Animal Learning & Behavior*, *23*, 197–206. https://doi.org/10.3758/BF03199935

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, PubMed: 27989778

Dubois, B., & Pillon, B. (1997). Cognitive deficits in Parkinson's disease. *Journal of Neurology*, *244*, 2–8. https://doi.org/10.1007/PL00007725, PubMed: 9007738

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*, 918–926. https://doi.org/10.1038/s41593-020-0658-y, PubMed: 32561883

Eckstein, M. K., & Collins, A. G. E. (2020). Computational evidence for hierarchically structured reinforcement learning in humans. *Proceedings of the National Academy of Sciences, U.S.A.*, *117*, 29381–29389. https://doi.org/10.1073/pnas.1912330117, PubMed: 33229518

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*, 1647–1659. https://doi.org/10.1093/cercor/bhu327, PubMed: 25601237

Failing, M. F., & Theeuwes, J. (2014). Exogenous visual orienting by reward. *Journal of Vision*, *14*, 6. https://doi.org/10.1167/14.5.6, PubMed: 24819737

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*, 48. https://doi.org/10.1186/s40337-021-00402-y, PubMed: 33865441

Frank, M. J. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks*, *19*, 1120–1136. https://doi.org/10.1016/j.neunet.2006.03.006, PubMed: 16945502

Frank, M. J., & Badre, D. (2012). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: Computational analysis. *Cerebral Cortex*, *22*, 509–526. https://doi.org/10.1093/cercor/bhr114, PubMed: 21693490

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*, 1030. https://doi.org/10.1038/s41467-021-21315-z, PubMed: 33589626

Fulvio, J. M., Akinnola, I., & Postle, B. R. (2021). Gender (Im)balance in citation practices in cognitive neuroscience. *Journal of Cognitive Neuroscience*, *33*, 3–7. https://doi.org/10.1162/jocn_a_01643, PubMed: 33078992

Gillan, C. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Voon, V., Apergis-Schoute, A. M., et al. (2014). Enhanced avoidance habits in obsessive–compulsive disorder. *Biological Psychiatry*, *75*, 631–638. https://doi.org/10.1016/j.biopsych.2013.02.002, PubMed: 23510580

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*, 20130475. https://doi.org/10.1098/rstb.2013.0475, PubMed: 25267818

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, PubMed: 30222107

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*, 2357–2371. https://doi.org/10.1162/jocn_a_01762, PubMed: 34272951

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory*, *70*, 119–136. https://doi.org/10.1006/nlme.1998.3843, PubMed: 9753592

Haber, S. N. (2011). Neuroanatomy of reward: A view from the ventral striatum. In J. A. Gottfried (Ed.), *Neurobiology of sensation and reward*. Boca Raton (FL): CRC Press/Taylor & Francis. 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*, 1252–1262. https://doi.org/10.1038/s41562-019-0725-0, PubMed: 31570762

Hickey, C., Kaiser, D., & Peelen, M. V. (2015). Reward guides attention to object categories in real-world scenes. *Journal of Experimental Psychology: General*, *144*, 264–273. https://doi.org/10.1037/a0038627, PubMed: 25559653

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, PubMed: 34222566

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*,

1399–1402. https://doi.org/10.1126/science.273.5280.1399, PubMed: 8703077

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). Chichester, UK: Wiley. 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, 762–774. https://doi.org/10.1002/hbm.21472, PubMed: 22113962

Krebs, R. M., & Woldorff, M. G. (2017). Cognitive control and reward. In T. Egner (Ed.), *The Wiley handbook of cognitive control* (pp. 422–439). Wiley. https://doi.org/10.1002/9781118920497.ch24

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press. https://doi.org/10.1016/B978-0-12-405888-0.00008-8

Lago, T., Davis, A., Grillon, C., & Ernst, M. (2017). Striatum on the anxiety map: Small detours into adolescence. *Brain Research*, 1654, 177–184. https://doi.org/10.1016/j.brainres.2016.06.006, PubMed: 27276526

Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, 40, 998–1009. https://doi.org/10.1002/ejsp.674

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, 505–513. https://doi.org/10.1038/nn2070, PubMed: 18327253

Lhermitte, F. (1983). 'Utilization behaviour' and its relation to lesions of the frontal lobes. *Brain*, 106, 237–255. https://doi.org/10.1093/brain/106.2.237, PubMed: 6850269

Libera, D. C., & Chelazzi, L. (2009). Learning to attend and to ignore is a matter of gains and losses. *Psychological Science*, 20, 778–784. https://doi.org/10.1111/j.1467-9280.2009.02360.x, PubMed: 19422618

Locke, H. S., & Braver, T. S. (2008). Motivational influences on cognitive control: Behavior, brain activation, and individual differences. *Cognitive, Affective, & Behavioral Neuroscience*, 8, 99–112. https://doi.org/10.3758/CABN.8.1.99, PubMed: 18405050

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, 231–247. https://doi.org/10.1093/cercor/bhab205, PubMed: 34231854

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, 1123–1136. https://doi.org/10.1098/rstb.2002.1099, PubMed: 12217179

Murty, V. P., Tompary, A., Adcock, A. R., & Davachi, L. (2017). Selectivity in postencoding connectivity with high-level visual cortex is associated with reward-motivated memory. *Journal of Neuroscience*, 37, 537–545. https://doi.org/10.1523/JNEUROSCI.4032-15.2016, PubMed: 28100737

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53, 139–154. https://doi.org/10.1016/j.jmp.2008.12.005

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, 191, 507–520. https://doi.org/10.1007/s00213-006-0502-4, PubMed: 17031711

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, 283–328. https://doi.org/10.1162/089976606775093909, PubMed: 16378516

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, PubMed: 19328775

Radulescu, A., Niv, Y., & Ballard, I. (2019). Holistic reinforcement learning: The role of structure and attention. *Trends in Cognitive Sciences*, 23, 278–292. https://doi.org/10.1016/j.tics.2019.01.010, PubMed: 30824227

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347–356. https://doi.org/10.1111/1467-9280.00067

Ribas-Fernandes, J. J. F., Solway, A., Diuk, C., McGuire, J. T., Barto, A. G., Niv, Y., et al. (2011). A neural signature of hierarchical reinforcement learning. *Neuron*, 71, 370–379. https://doi.org/10.1016/j.neuron.2011.05.042, PubMed: 21791294

Rmus, M., McDougle, 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, PubMed: 35194556

Roelfsema, P. R., van Ooyen, A., & Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. *Trends in Cognitive Sciences*, 14, 64–71. https://doi.org/10.1016/j.tics.2009.11.005, PubMed: 20060771

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *Journal of Neuroscience*, 29, 7191–7198. https://doi.org/10.1523/JNEUROSCI.0979-09.2009, PubMed: 19494141

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, 79, 217–2140. https://doi.org/10.1016/j.neuron.2013.07.007, PubMed: 23889930

Solley, C. M., & Murphy, G. (1960). Effects of practice and reward. In C. M. Solley & G. Murphy (Eds.), *Development of the perceptual world* (pp. 81–103). New York, NY: Basic Books/Hachette Book Group. https://doi.org/10.1037/11120-005

Son, L. K., & Sethi, R. (2006). Metacognitive control and optimal learning. *Cognitive Science*, 30, 759–774. https://doi.org/10.1207/s15516709cog0000_74, PubMed: 21702835

Steinglass, J., & Walsh, B. T. (2006). Habit learning and anorexia nervosa: A cognitive neuroscience hypothesis. *International Journal of Eating Disorders*, 39, 267–275. https://doi.org/10.1002/eat.20244, PubMed: 16523472

Sternberg, D. A., & McClelland, J. L. (2012). Two mechanisms of human contingency learning. *Psychological Science*, 23, 59–68. https://doi.org/10.1177/0956797611429577, PubMed: 22198929

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, PubMed: 28515705

Swanson, S. D., & Tricomi, E. (2014). Goals and task difficulty expectations modulate striatal responses to feedback. *Cognitive, Affective, & Behavioral Neuroscience*, 14, 610–620. https://doi.org/10.3758/s13415-014-0269-8, PubMed: 24638235

Tajima, S., Drugowitsch, J., & Pouget, A. (2016). Optimal policy for value-based decision-making. *Nature Communications*, 7, 12400. https://doi.org/10.1038/ncomms12400, PubMed: 27535638

Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience*, 29, 2225–2232. https://doi.org/10.1111/j.1460-9568.2009.06796.x, PubMed: 19490086

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, *59*, S251–S278. https://doi.org/10.1086/296365

Vandaele, Y., & Ahmed, S. H. (2021). Habit, choice, and addiction. *Neuropsychopharmacology*, *46*, 689–698. https://doi.org/10.1038/s41386-020-00899-y, PubMed: 33168946

Versace, F., Robinson, J. D., Lam, C. Y., Minnix, J. A., Brown, V. L., Carter, B. L., et al. (2010). Cigarette cues capture smokers' attention: Evidence from event-related potentials. *Psychophysiology*, *47*, 435–441. https://doi.org/10.1111/j.1469-8986.2009.00946.x, PubMed: 20070579

Voon, V., Baek, K., Enander, J., Worbe, Y., Morris, L. S., Harrison, N. A., et al. (2015). Motivation and value influences in the relative balance of goal-directed and habitual behaviours in obsessive–compulsive disorder. *Translational Psychiatry*, *5*, e670. https://doi.org/10.1038/tp.2015.165, PubMed: 26529423

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*, 953–956. https://doi.org/10.1038/35082081, PubMed: 11418860

Waskom, M. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*, 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. *Journal of Neuroscience*, *34*, 10743–10755. https://doi.org/10.1523/JNEUROSCI.5282-13.2014, PubMed: 25100605

Waskom, M. L., Okazawa, G., & Kiani, R. (2019). Designing and interpreting psychophysical investigations of cognition.

Neuron, *104*, 100–112. https://doi.org/10.1016/j.neuron.2019.09.016, PubMed: 31600507

Waskom, M. L., & Wagner, A. D. (2017). Distributed representation of context by intrinsic subnetworks in prefrontal cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *114*, 2030–2035. https://doi.org/10.1073/pnas.1615269114, PubMed: 28174269

Weintraub, D. (2008). Dopamine and impulse control disorders in Parkinson's disease. *Annals of Neurology*, *64 (Suppl. 2)*, S93–S100. https://doi.org/10.1002/ana.21454, PubMed: 19127573

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, PubMed: 23935581

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, PubMed: 31769410

Wilson, R. C., Shenhav, A., Straccia, M., & Cohen, J. D. (2018). The eighty five percent rule for optimal learning. *bioRxiv*, 255182. https://doi.org/10.1101/255182

Wood, W. (2017). Habit in personality and social psychology. *Personality and Social Psychology Review*, *21*, 389–403. https://doi.org/10.1177/1088868317720362, PubMed: 28737111

Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, *67*, 289–314. https://doi.org/10.1146/annurev-psych-122414-033417, PubMed: 26361052

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*, 464–476. https://doi.org/10.1038/nrn1919, PubMed: 16715055