

Journal of Psychopathology and Clinical Science

The Challenge of Learning Adaptive Mental Behavior

Peter F. Hitchcock and Michael J. Frank

Online First Publication, May 30, 2024. <https://dx.doi.org/10.1037/abn0000924>

CITATION

Hitchcock, P. F., & Frank, M. J. (2024). The challenge of learning adaptive mental behavior.. *Journal of Psychopathology and Clinical Science*. Advance online publication. <https://dx.doi.org/10.1037/abn0000924>

The Challenge of Learning Adaptive Mental Behavior

Peter F. Hitchcock^{1, 2} and Michael J. Frank^{2, 3}

¹Department of Psychology, Emory University

²Cognitive, Linguistic, and Psychological Sciences, Brown University

³Carney Institute for Brain Science, Brown University

Many psychotherapies aim to help people replace maladaptive mental behaviors (such as those leading to unproductive worry) with more adaptive ones (such as those leading to active problem solving). Yet, little is known empirically about how challenging it is to learn adaptive mental behaviors. Mental behaviors entail taking mental operations and thus may be more challenging to perform than motor actions; this challenge may enhance or impair learning. In particular, challenge when learning is often desirable because it improves retention. Yet, it is also plausible that the necessity of carrying out mental operations interferes with learning the expected values of mental actions by impeding credit assignment: the process of updating an action's value after reinforcement. Then, it may be more challenging not only to perform—but also to learn the consequences of—mental (vs. motor) behaviors. We designed a task to assess learning to take adaptive mental versus motor actions via matched probabilistic feedback. In two experiments ($N = 300$), most participants found it more difficult to learn to select optimal mental (vs. motor) actions, as evident in worse accuracy not only in a learning but also test (retention) phase. Computational modeling traced this impairment to an indicator of worse credit assignment (impaired construction and maintenance of expected values) when learning mental actions, accounting for worse accuracy in the learning and retention phases. The results suggest that people have particular difficulty learning adaptive mental behavior and pave the way for novel interventions to scaffold credit assignment and promote adaptive thinking.

General Scientific Summary

We are often asked to think harder, remain positive, and not sweat the small stuff. But why is it so challenging to ingrain adaptive mental behaviors, such as those that lead to thinking in healthy and productive ways? Using a novel task that directly compared the ability to learn optimal mental versus motor behaviors via trial and error, we found that (overall) people had more difficulty learning optimal mental behaviors and traced this difficulty to the formation of less robust expected values for mental actions.

Keywords: behavior modification, computational modeling, meta-control, reinforcement-learning, perseverative thinking


Supplemental materials: <https://doi.org/10.1037/abn0000924.supp>

We are often asked to think harder, remain positive, and not sweat the small stuff. But why is it so challenging to ingrain adaptive mental behaviors, such as those that lead to thinking in healthy and productive ways? This article introduces a novel task to investigate the

comparative difficulty of learning to select adaptive mental (cognitive) as opposed to motor (overt) actions.

This question is important because a number of psychotherapies take a behavioral approach to cognition. That is, they assume that

Alexander J. Shackman served as action editor.

Peter F. Hitchcock  <https://orcid.org/0000-0001-7606-5132>

Peter F. Hitchcock was supported by the National Institute of Mental Health Grant F32MH123055. Michael J. Frank was supported by the National Institute of Mental Health Grants P50MH119467 and R01MH084840-08A and the Office of Naval Research Multidisciplinary University Research Initiatives Award N00014-23-1-2792. The authors thank the Frank Lab at Brown University; the Neuroscience and Mental Health Group at the Institute of Cognitive Neuroscience at University College London; and Rex Liu, Ryan Smith, and Brianna Yamasaki for helpful comments that contributed to the design, analyses, and/or discussion. All participants gave informed consent before the study, and the study was approved by the Brown University Institutional Review Board. This work has been preprinted at <https://psyarxiv.com/agprs>. The code used to

produce the results and figures is available at https://github.com/peter-hitchcock/cog-acts_analysis. Task data will be posted to the following open repository upon publication: <https://nivlab.github.io/pendata>. This study was not preregistered.

Peter F. Hitchcock served as lead for data curation, formal analysis, funding acquisition, investigation, methodology, project administration, software, validation, visualization, writing—original draft, and writing—review and editing and contributed equally to resources. Michael J. Frank served as lead for supervision and served in a supporting role for formal analysis, funding acquisition, investigation, methodology, validation, visualization, writing—original draft, and writing—review and editing. Peter F. Hitchcock and Michael J. Frank contributed equally to conceptualization.

Correspondence concerning this article should be addressed to Peter F. Hitchcock, Department of Psychology, Emory University, 36 Eagle Row, Atlanta, GA 30322, United States. Email: peter.hitchcock@emory.edu

both cognitive and overt activities are types of behavior; cognitive behaviors just happen to take place “between the ears” rather than in the external world. Hence, the therapies apply behavior-modification strategies aimed at uncovering the function of maladaptive mental behaviors and finding adaptive alternatives that fulfill a similar function and can be practiced in their place (e.g., S. A. Hayes et al., 2010; S. C. Hayes et al., 2011; Martell et al., 2021; Watkins, 2018).

Behavior modification assumes that the consequences of behavior can be learned—in particular, that, after sufficient experience, adaptive behaviors will naturally replace maladaptive ones within an individual’s repertoire (Kazdin, 2012; Ramnero & Törneke, 2008). Yet, little is known empirically about the comparative difficulty of learning adaptive cognitive versus overt behaviors.¹

To investigate this question, we designed a novel bandit reinforcement-learning (RL) task. Bandit tasks are workhorses of RL research that have been extensively investigated in human and non-human animals (e.g., Frank et al., 2004; Pessiglione et al., 2006; Schoenbaum et al., 2003). In our task, the baseline “overt” condition is similar to standard bandit tasks that simply require the learner to acquire stimulus–response associations, where the response is a motor action. We compared this to a “cognitive” bandit with matched probabilistic reinforcement contingencies and motor demands, but where the required responses were cognitive actions rather than simply overt actions (see below).²

Our focus is not on the performance of the actions themselves, but on the ability to learn which action to select based on reinforcement history. One might naturally expect that the increased challenge associated with performing a cognitive operation, which is entailed by taking a cognitive action, would impede learning the cognitive action’s value. Yet, in many cases, the opposite is true: greater challenge during learning actually enhances later retention. For instance, retrieving information rather than simply restudying it improves later retention (Karpicke & Roediger, 2008); this is just one of a number of so-called “desirable difficulties” in learning (Bjork & Bjork, 2011). Notably, in RL tasks, greater demands on working memory while learning lead to slower initial acquisition of stimulus–response contingencies but paradoxically better retention of these contingencies (via enhanced retention of their expected values) in a later test phase (Collins et al., 2014, 2017; Rac-Lubashevsky et al., 2023). This enhanced retention appears to be mediated by greater RL-based neural signaling under high working-memory load (Collins & Frank, 2018; Rac-Lubashevsky et al., 2023). Hence, greater difficulty in the form of higher working-memory demand appears to increase activity in the RL system that ultimately enhances retention. Thus, a crucial feature of our experimental design was the inclusion of a test phase to assess retention.

We considered two possibilities: First, If cognitive (vs. overt) actions are more challenging, initial learning might be slowed but retention of learned values might be enhanced. Conversely, we hypothesized that the need to perform a cognitive operation, which is entailed by taking a cognitive action, would interfere with RL by disrupting credit assignment: the process of updating the cognitive action’s value after reinforcement (Sutton & Barto, 2018). Notably, whereas working-memory load increases prediction errors in the overt stimulus–response tasks described above (Collins & Frank, 2018; Collins et al., 2014; Rac-Lubashevsky et al., 2023), these errors presumably can be correctly attributed to the stimulus and action just selected—and this correct attribution is why retention is ultimately enhanced. In contrast, the performance of a cognitive operation might disrupt the ability to attribute prediction errors to the cognitive

action itself. Broadly consistent with this, prior work suggests that people ascribe credit not only to outcome-relevant aspects of their actions, but also incorrectly to irrelevant aspects that are entailed by the action (such as spatiomotor aspects orthogonal to reward; Shahar et al., 2019; see also Jocham et al., 2016; Lamba et al., 2023).

Another reason to suspect that credit assignment may be disrupted for cognitive actions is that, biologically, credit assignment depends on a delicate orchestration whereby dopaminergic signals are (or are not) propagated to subregions representing aspects of one’s action (and state) responsible for the outcome (Hamid et al., 2021). It is quite plausible that this delicate arrangement is made more challenging when a cognitive operation is interposed between an action’s initiation and its outcome—as is necessarily the case for cognitive actions.

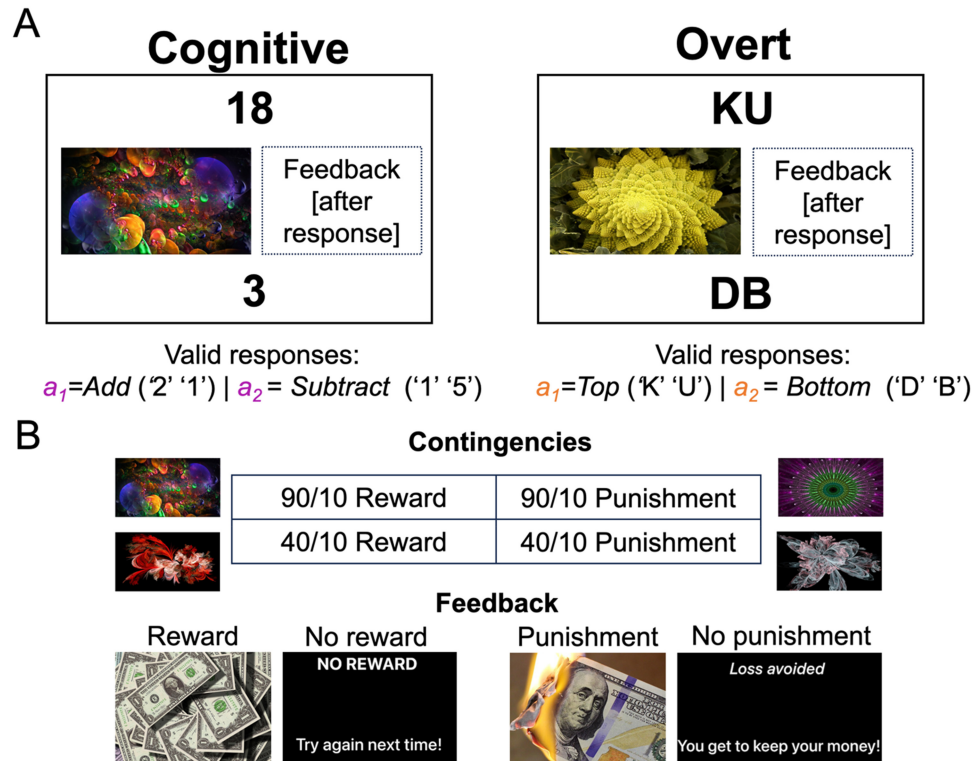
It is also noteworthy that working memory is often required for credit assignment. In particular, it facilitates learning the expected values of specific actions (Asaad et al., 2017; Frank & Claus, 2006; Gold et al., 2012; Hernaus et al., 2018, 2019). Yet, the cognitive operations entailed by performing cognitive actions also tend to rely on working memory. Thus, the dual demands of performing a cognitive operation, on the one hand, and assigning credit to the abstract action itself, on the other hand, may draw on shared working-memory resources. This may lead to particular difficulties in acquiring expected values for cognitive actions (i.e., estimates of their long-run average consequences in different situations)—as has been found, for example, in a neural-network model wherein credit assignment to cognitive actions was especially difficult when multiple working-memory representations were held in mind (O’Reilly & Frank, 2006).

We developed a novel cognitive actions task (Figure 1) to investigate the distinct possibilities that adaptive learning would be enhanced versus impaired for cognitive (vs. overt) behaviors. The design was such that specific sequences (e.g., “18 + 3 = 21” in the cognitive condition or keys “K” and “U” in the overt condition) changed across trials, whereas high-level cognitive versus overt actions had fixed reward or punishment probabilities in the same state. Thus, we were able to

¹ The notion that there may be particular challenges involved in modifying cognitive vs. overt behavior is well established. For instance, Kazdin (2012, pp. 3–4) describes practical reasons that psychotherapies often focus on changing overt rather than cognitive behavior. Hayes and colleagues have considered particularities of verbal behavior that might make them especially challenging to modify (reviewed in S. C. Hayes et al., 2011, pp. 39–52). Yet, our study is the first, to our knowledge, to directly compare the ability to learn adaptive cognitive versus overt behaviors within a reinforcement-learning design.

² Of note, our labels of the two conditions as “overt” and “cognitive” are heuristic rather than strict distinctions, as we assume both conditions require cognition, for instance, to maintain the task set and remember responses across delays (see Collins & Frank, 2012 for evidence that even quite simple RL tasks involve working memory). Nonetheless, as described in the text, our conditions are distinguished by the fact that—orthogonal to these baseline demands present in both conditions—the cognitive condition alone requires performing a cognitive operation (e.g., add or subtract) to take the action on each trial, and it is the value of this abstract cognitive operation itself that must be learned. It is this distinction that we are highlighting with the labels cognitive versus overt. Our heuristic labeling strategy is similar to that employed in past reinforcement-learning tasks, such as a task that manipulated model-based versus model-free contributions (Daw et al., 2011), which were labeled as such although orthogonal model-based representational requirements are present even in the model-free case (see Shahar et al., 2019); and in a reinforcement-learning and working memory task that manipulated working memory load, even though some working memory contribution is presumed to be needed even in the low WM conditions (Collins & Frank, 2012, 2018; Rac-Lubashevsky et al., 2023).

Figure 1
The Cognitive Actions Task (Experiment 1 Variant)



Note. (A) In the cognitive actions task, participants must learn through trial and error to take the best of two actions in various states signaled by different fractal images (on the left of each screen). In the cognitive condition, the actions are add or subtract. In the example shown, the participant is presented with “18” and “3” (after 500 ms of just the fractal being displayed). Thus, they can type “21” to select add or “15” to select subtract (the numbers to add/subtract change every trial, and participants type the keys to enter the sum/difference—e.g., typing “2” then “1” for add on this trial). After their response, reward or punishment feedback is delivered (in the learning period only; in the test period, the task is the same but no feedback is shown). In the overt condition, the actions are top or bottom; in the example shown, the participant would type “K” + “U” for top or “D” + “B” for bottom (the letters change every trial) and then receive reinforcement. (B) In both conditions, there are four different states (equivalently, bandits) each signaled by different fractal images, where each state/fractal is associated with a different contingency. The contingencies in the states were created by crossing the feedback percentages {90-10, 40-10} with valence {Reward, Punishment}. For example, in {90-10 Reward}, reward is delivered 90% of the time that the optimal action (e.g., add) is selected and 10% of the time that the nonoptimal action (e.g., subtract) is selected (the optimal actions in different contingencies were counterbalanced across participants). The “Feedback” heading shows all types of feedback when reward or punishment was/was not received. Because contingencies and motor demands were matched, the key difference between the conditions was that the cognitive condition alone required performing a mental operation. In Experiment 2, we employed a second variant of the cognitive actions task that involved different cognitive actions, which is shown in Figure 1 in the online supplemental materials. The contingencies and feedback were the same as in Experiment 1. See the online article for the color version of this figure.

isolate the ability to learn optimal cognitive (e.g., add) versus overt (e.g., top) actions themselves (Figure 1).

In two experiments with different variants of the task ($N = 300$), we found that, overall, people had more difficulty learning which cognitive actions were optimal, compared to the baseline overt-action learning task. Crucially, this impairment persisted in a later test (retention) phase. Computational modeling traced the impairment to less ability to construct and retain expected values for cognitive actions, consistent with impaired credit assignment (see “Discussion” regarding open questions that this study raises about the specific mechanisms responsible for the impairment). Of

note, despite an overall group effect in both experiments, there was substantial heterogeneity—with a subset of participants actually showing better accuracy (and retention of expected values) in the cognitive condition. Nevertheless, that adaptive cognitive-action learning was more difficult for most people suggests that it is not comparably easy as overt-action learning. As such, it may be beneficial for psychotherapies that target mental behaviors to develop extra support to scaffold learning about them. Moreover, the heterogeneity that we observed paves the way for future research on individual differences. Ultimately, such research may facilitate targeted interventions that provide extra scaffolding or adjunctive

treatment (e.g., neuromodulation) for those who require extra help learning the consequences of their mental behavior.

Method

Participants

Across both Experiments 1 and 2, participants from the United States ages 18–65 ($N = 300$; $n = 150$ in each study) were recruited via Prolific and provided informed consent via a consent form approved by the Brown University Institutional Review Board. Participants received baseline compensation (\$9.50 USD/hr) and a bonus based on accuracy (up to \$3; they were told that we expected about 1/3 of participants to attain \$3 and were subsequently compensated \$1/2/3 depending on the tercile of their accuracy, evaluated in terms of proportion correct during the learning phase).

In each experiment, participants completed the cognitive actions task followed by questionnaires. In quality control checks similar to those in prior work (Hitchcock et al., 2022; Radulescu et al., 2016), we excluded participants who in the learning phase (a) performed indistinguishably from chance in the 90-10 probability contingencies ($<55.93\%$ correct, which is the binomial mean + 1.5 SD for $p = .5$, $n = 160$ trials) and/or (b) responded with the same action (e.g., add) for any 50+ consecutive-trial streak. After removals, the sample sizes were as follows: $n = 125$ in Experiment 1, 45.6% female, 54.4% male; M_{age} (SD) = 36.65 (11.80); ethnic identity from Prolific profile: 5.60% Asian, 6.40% Black; 4.00% Mixed; 2.4% Other; 80% White; 1.60% data not available (data expired) 1.6%, and $n = 138$ in Experiment 2, 53.33% female, 46.67% male; M_{age} (SD) = 37.90 (11.28); ethnic identity from Prolific profile: 2.90% Asian, 3.62% Black; 9.42% Mixed; 3.62% Other; 76.10% White; 4.35% data not available (data expired); the Experiment 2 demographics do not include three participants who did not have sex and age in their Prolific profile).

Tasks

The variants of the cognitive actions task used in the first and second experiments had the same structure. On each trial, participants first saw a fractal image, signaling which state they were in, for 500 ms. This image remained on the screen while stimuli appeared at the top and bottom of the screen that allowed them to take one of two actions (in RL terminology, each state was a 2-armed bandit). Reinforcement was immediately delivered following the action while the stimuli and fractal remained on screen (Figure 1 and Figure 1 in the online supplemental materials).

All participants completed cognitive and overt conditions. The key difference between conditions was that only the former required performing a cognitive action. In the Experiment 1 variant of the task, the cognitive actions were adding/subtracting two numbers and the overt actions involved selecting letters at the top/bottom of the screen (Figure 1). Accuracy in the task was evaluated as ability to learn to select the more optimal (most rewarding or least punishing) action—for instance, in Experiment 1, whether subtract (vs. add) was the optimal action in a given state (fractal image), which had to be learned through repeated trials in that state with changing numbers (e.g., on one trial the numbers might be “15” and “4” and the response needed to select subtract would be “11”; in the next, “18” and “2” and the response needed would be “16”). Note that accuracy is distinct from response validity (for instance, for “15” and “4” the two valid responses

are “19” for add or “11” for subtract; see below for further details on how invalid responses were handled). In the Experiment 2 variant of the task, the cognitive actions were alphabetizing/reverse-alphabetizing letters and the overt actions were selecting numbers on the diagonal/reverse-diagonal of the screen (Figure 1 in the online supplemental materials).

Participants completed two practice phases. In the first, they practiced performing each action in each condition (i.e., in Experiment 1, typing the keys corresponding to add, subtract, top, and bottom) with no fractal image yet displayed. Next, they practiced a simplified version of the task with two states (signaled by different fractals) for which opposite actions were deterministically correct, with reinforcement displayed to signify that the action was correct (for example, in the Experiment 1 cognitive condition, add was the correct (rewarded/loss avoiding) response for one image and subtract was the correct response for the other). Participants were required to enter four consecutive correct responses in practice phase 1 and at least five out of six correct responses in practice phase 2 to continue in the task.

Participants next completed the learning phase of the task. Here, they attempted to learn the optimal action for each state through probabilistic reinforcement feedback. The learning phase comprised four blocks of 80 trials each of alternating conditions, with first-condition order counterbalanced (i.e., {overt, cognitive, overt, cognitive} or {cognitive, overt, cognitive, overt}). Each block comprised 20 trials each of four states (signaled by different fractal images) specific to the condition, with trial order randomized. The contingencies for the four states, which were constant throughout the task in each condition, were constructed by crossing valence {reward, punishment} and outcome probability {90-10, 40-10}, leading to {90-10 Reward, 40-10 Reward, 90-10 Punishment, 40-10 Punishment} bandits. The images displayed to signal reward versus no-reward were a picture of money (see Figure 1) or the text “NO REWARD | Try again next time!” respectively; and for punishment versus no-punishment, were a picture of burning money or the text “Loss avoided | You get to keep your money!” respectively.

Participants next completed two additional phases (order counterbalanced). A test phase comprised trials identical to the learning phase but without feedback, thus providing a measure of asymptotic retention; participants performed eight trials of each of the four bandits in the two conditions (32 trials/condition = 64 trials). A stimulus-valuation phase required choosing between the images (states) from the learning phase, now presented in pairs. Participants were instructed to “Select the picture that ‘feels’ like it will win the most,” by which we sought to convey that they should pick the state that had been most rewarded (least punished). There were two subphases of the stimulus-valuation phase: one involving choosing between states from the same condition (overt–overt and cognitive–cognitive) and the other states from different conditions (overt–cognitive). In the former, because each state from the learning phase corresponded to a unique contingency, all pairs differed in their expected value (e.g., 90-10 Reward vs. 40-10 Reward); participants completed four repetitions of the six possible state pairs in the two conditions (48 trials/participant). In the latter, pairings were limited to states with the same expected value (e.g., 90-10 Reward overt vs. 90-10 Reward cognitive); participants completed eight repetitions of the four bandit pairs (32 trials/participant). Participants selected images on the right or left of the screen by pressing the corresponding arrow keys.

Please see the online supplemental materials for further information about the task parameters.

Posttask Questionnaires

Following the tasks, participants completed individual difference questionnaires, including the 15-item Perseverative Thinking Questionnaire (Ehring et al., 2011) and 10-item Rumination-Response Style questionnaire (Trenor et al., 2003), to permit examination of whether individual differences in thinking style correlated with model-based differences, and the 24-item Behavioral Inhibition/Behavioral Activation scales (Carver & White, 1994), to examine if individual differences in self-reported aversive and appetitive motivation respectively correlated with differences in reward or punishment learning irrespective of condition.

Code, Data Availability, and Task Coding

The code used to produce the results and figures is available at https://github.com/peter-hitchcock/cog-acts_analysis. Task data will be posted to the following open repository upon publication: <https://nivlab.github.io/opendata>. The task was coded in Honeycomb (Provenza et al., 2022).

Analyses

Statistical Data Analyses

Data were preprocessed and analyzed using R (Version 4.1.2; R Core Team, 2021). For analyzing the effects of condition (overt vs. cognitive) and delay (i.e., whether or not the same bandit was played on consecutive trials) on accuracy (correct: yes/no) in the learning and test phases, and the effect of reward history on choice in the stimulus-valuation phase, we built mixed-effects logistic regression models using the lme4 package (Bates et al., 2014) with p -values estimated using Satterthwaite's method for approximating degrees of freedom. The variance inflation factor for all multivariate regression models was <1.1 , suggesting no issues with collinearity. Tables 2–4 in the online supplemental materials report full model specifications, variable coding (contrast and z-scoring), and statistics for the regression models.

The stimulus-valuation phase examined selection between state (i.e., fractal stimulus) pairs, which had been learned about separately during the learning phase. As described above, participants were instructed to pick the state that had led to the best outcomes during the training phase ("Method"). Because these selections were at the level of the entire stimulus, we examined choice as a function of its (model-agnostic) reward history, rather than (model-derived) Q-values that were specific to the different actions in the state.

The computational-model parameter, ϕ , that was allowed to vary across conditions in the most successful computational model was heavily left skewed in both conditions (see Table 7 in the online supplemental materials). Parameter recovery for the sign of the difference in this parameter was also relatively good (71.43% correct sign) whereas recovery for the difference in values was weaker (Spearman's ρ : $\phi^{\text{cog}} - \phi^{\text{overt}} = .56$). Thus we used the nonparametric χ^2 test to examine the frequency with which this parameter was higher in one condition than the other.

Computational Modeling Analyses

To model learning and choice dynamics in the task, we fit various models that acquired values for the two actions in each state

(signaled by fractal images) via RL, Bayesian, and hybrid Bayesian–RL mechanisms. A full description of the computational models evaluated is in the online supplemental materials.

Model Fitting Procedure. We estimated model parameters using an empirical Bayes approach, wherein model parameters were estimated for individual subjects using maximum likelihood estimation and then in a second step these estimates were regularized by using group-level statistics as a prior on the individual estimates, thereby shrinking each parameter estimate toward the group-level mean (Casella, 1985; Piray et al., 2019). This hierarchical approach, whereby subject-level estimates are constrained based on the assumption that parameters are drawn from a population distribution estimated via the group-level statistics, decreases overfitting and thus tends to improve parameter estimation (Katahira, 2016).

Specifically, the maximum likelihood procedure followed in the first step minimized the negative log likelihood for each participant's observed choices summed over all trials via the "solnp" function in the "Rsolnp" package in R (Ghalanos & Theussl, 2011). Once all subjects had been run, a penalty—the multivariate normal density of group-level parameter means and covariance matrix—was constructed, and all subjects were re-fit with the negative log of this penalty added to the negative log likelihood, thereby constraining estimates in the optimization according to the group-level statistics. This approach is an approximation to the expectation-maximization algorithm (which iterates between group and subject-level parameters; e.g., Guitart-Masip et al., 2012) and tends to improve parameter recovery relative to maximum likelihood estimation without such regularization (Frey et al., 2021), including for the current task/model.

To avoid local minima, we ran optimization 50 times for each subject on the first step (used for estimating the mean and covariance matrix for the second step) and then 20 times at the second step (used to derive the finalized penalized subject-level estimates). This procedure was repeated twice for each subject and the estimate with the lowest negative log likelihood across these two runs was used.

Model Comparison Procedure. We used the Akaike information criterion (AIC), which is $2 \times (\text{negative log likelihood} + \text{the number of free parameters})$, as a metric of model fit that penalizes for model complexity in terms of the number of free parameters (Akaike, 1998).

Model Validation Procedure. Model validation was performed by taking our subject-level parameter estimates derived via the fitting procedure described above and simulating task data, using the same contingencies as experienced by participants but simulating actions and rewards rather than relying on the empirical ones. These data were compared to various key features of the data to examine how well the model was able to capture these aspects.

Parameter Recovery Procedure. We generated parameters from a truncated multivariate gaussian distribution concentrated around the median of the empirical estimates, simulated data based on these parameters as just described, and then attempted to recover the true values following the empirical Bayes procedure described above in "Model Fitting Procedure." The simulated versus recovered values were compared using Spearman's ρ .

Results

Participants were recruited via Prolific (see "Method") for two experiments that manipulated demands on RL by comparing a standard bandit RL condition, in which participants learned to select

overt motor actions that were differentially reinforced, to a condition with matched probabilistic contingencies in which the actions were cognitive. In Experiment 1 ($N = 150$) participants performed the version of the task described in Figure 1. Here, in the cognitive condition participants learned whether to add or subtract two numbers. They were not told whether to add or subtract, but instead had to learn over trials whether the cognitive action of adding or subtracting was probabilistically more often reinforced in different states. In the overt condition, the contingencies were the same, but this condition only required learning the value of overt actions: typing letter pairs on the top or bottom of the screen (note that the motor demands are matched to the cognitive condition).

A difficulty with attributing any differences observed with the Experiment 1 variant to cognitive (vs. overt) actions per se, is that such differences could also arise because the cognitive condition uniquely required numerical cognition. Thus, in Experiment 2 ($N = 150$) participants performed an alternate version of the task (Figure 1 in the online supplemental materials) in which the overt condition now involved numbers (typing numbers on the slash-diagonal or backslash-diagonal of the screen), whereas the cognitive condition involved performing a mental operation on letters (alphabetizing vs. reverse-alphabetizing). The use of both variants allowed us to isolate how performing a mental operation, which is entailed by taking a cognitive action (i.e., mental behavior), influences learning—while crossing the demands for lexical versus numerical cognition across the experiments.

In both experiments, the cognitive and overt conditions had the same probability and valence contingencies and matched motor demands. The contingencies were unknown to participants, who therefore had to learn to select the best action in different states (signaled by different fractal images) through trial and error. Participants performed a learning phase with 40 trials per contingency in each of the four contingencies per condition (160 total trials per condition; 320 learning trials overall) followed by a test phase, which was identical to the learning phase but with feedback withheld and which therefore allowed an assessment of retention (“Method”). Participants also performed a state-valuation phase that involved selecting between novel fractal image pairs to assess the subjective value they attributed to whole states while averaging over actions (see “Method” and “Supplemental Results” in the online supplemental materials).

Accuracy Was Lower in the Cognitive (Vs. Overt) Condition, Albeit With Substantial Heterogeneity Across Participants

As hypothesized, in both experiments, accuracy in choosing the most rewarding (least punishing) action was worse in the cognitive compared to overt condition in both the learning and test phases (Figure 2); Experiment 1: learning, condition β (SE) = $-.18$ (0.07), $p < .008$; test, condition β (SE) = $-.46$ (0.15), $p < .003$; Experiment 2: learning, condition β (SE) = $-.17$ (0.07), $p < .02$; test, condition β (SE) = $-.35$ (0.16), $p < .03$. Results were largely consistent across various robustness checks (“Supplemental Results” in the online supplemental materials; see “Method” for the description of regression models and Tables 2–4 in the online supplemental materials for their full specification and results).

Although there was a group effect reflecting worse accuracy in the cognitive (than overt) condition, there was substantial heterogeneity

among participants—a substantial subset actually had higher accuracy in the cognitive than overt condition (Experiment 1: learning, 37.6% better in cognitive, with 2.4% showing no difference; test, 32.8% better in cognitive, with 10.4% showing no difference; Experiment 2: learning, 39.13% better in cognitive, with <1% showing no difference; test, 35.50% better in cognitive, with 18.84% showing no difference).

Accuracy was modestly correlated between conditions, suggesting the presence of individual differences that led to differential accuracy in this learning task irrespective of condition (Experiment 1: learning, $r = .44$, $p < 5e-7$; test, $r = .45$, $p < 5e-7$; Experiment 2: learning, $r = .46$, $p < 5e-8$; test, $r = .38$, $p < 5e-6$).

Accuracy Declined Due to Delay but Without Consistent Moderation by Condition

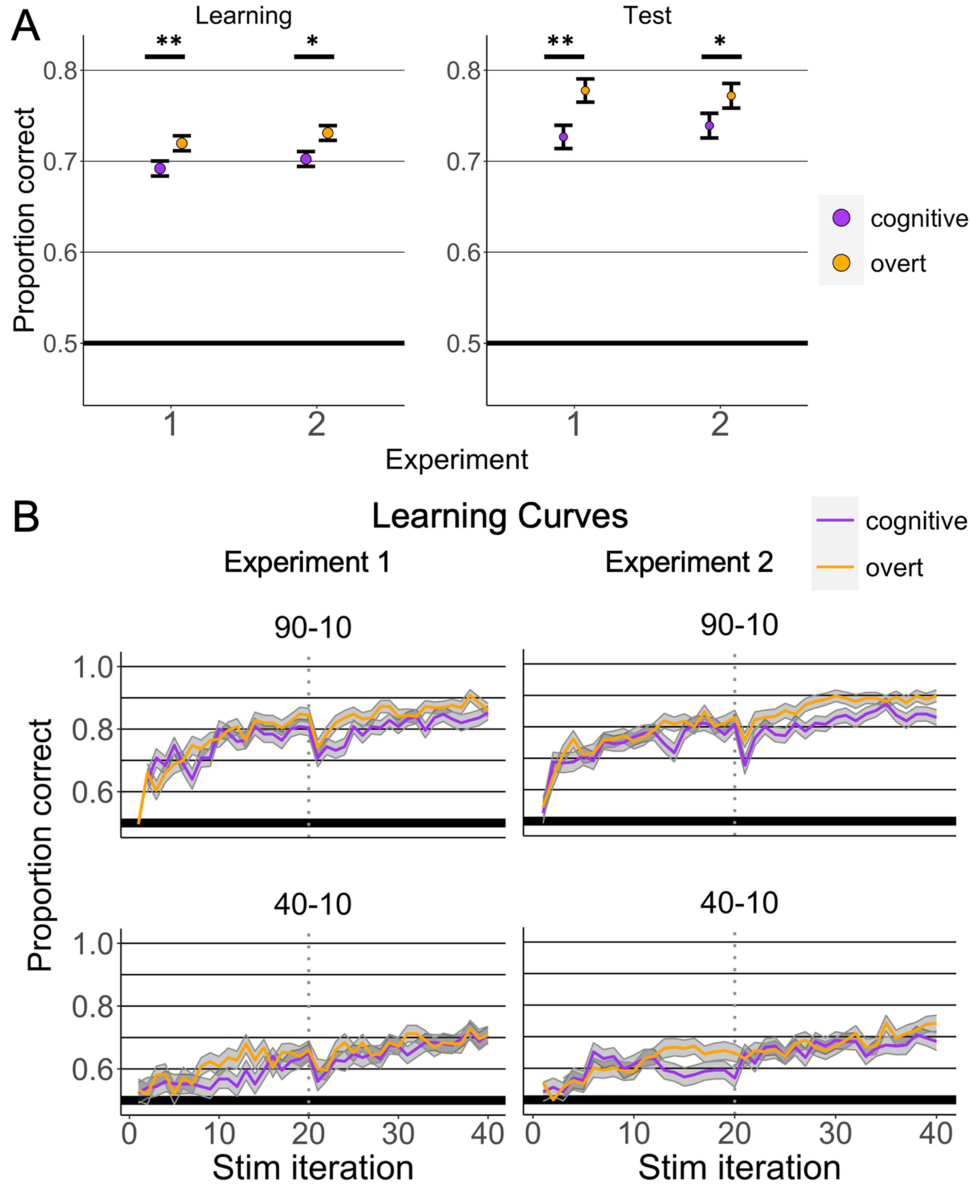
We next asked if accuracy during learning decreased when there was a delay between states (fractal stimuli)—that is, whether states (bandits) were repeated on consecutive trials or rather interrupted by other bandits. Deleterious effects of delay on accuracy are common in learning tasks with working-memory demands (e.g., Collins et al., 2017), hence the greater working-memory demands in the cognitive condition might lead to an especially strong delay effect in this condition (although note that this is not the only role of working memory in learning, as it also appears to play a role in constructing persistent expected action values, e.g., Gold et al., 2012). Visually, in both experiments, there was a clear decremental effect of delay on accuracy in both conditions (Figure 2 in the online supplemental materials) and indeed delay predicted decreased accuracy statistically, Experiment 1: delay β (SE) = $-.36$ (0.03), $p < 2e-16$; Experiment 2: delay β (SE) = $-.31$ (0.03), $p < 2e-16$. However, the effect of delay was only stronger in the cognitive (than overt) condition in the second experiment, Experiment 1: Delay \times Condition β (SE) = $-.01$ (0.06), ns ; Experiment 2: Delay \times Condition β (SE) = $-.12$ (0.06), $p < .04$, and no significant interaction was present in either experiment when using continuous rather than binary delay ($ps > .34$). Thus, the results show delay effects during learning irrespective of condition, but without consistent evidence for a stronger tax of delay on the cognitive condition.

Notably, the results from the previous section—which showed that accuracy was lower in the cognitive (vs. overt) condition not only in the learning phase, but also in the test phase where feedback is no longer provided on each trial—provide converging evidence that the accuracy differences between conditions are not solely driven by the ability to retain contingencies explicitly during learning, but also by weaker retention of learned associations asymptotically—which also may depend on working memory (e.g., Gold et al., 2012).

Although the findings thus far provide support for our overall hypothesis that learning to select cognitive (vs. overt) actions is more difficult—not only during initial learning but also during a test (retention) phase—they do not identify specific learning and choice mechanisms responsible for this impairment. Thus, we next constructed formal trial-wise learning models.

An RL Computational Model Was Able to Capture Key Task Effects

We compared a variety of computational models that might plausibly capture learning and choice processes in the task in terms of

Figure 2*Accuracy Differences Between Conditions*

Note. (A) Proportion of correct choices (i.e., selecting the optimal action) in Experiments 1 and 2 in the cognitive versus overt conditions in the learning (left) and test (right) phases. Colored points indicate means and error bars show ± 1 within-subject SEM. (B) Learning curves (i.e., proportion of time choosing the optimal action as a function of exposure to each fractal/state, denoted by stim iteration, during the learning phase) shown separated into easier (90-10) and more difficult (40-10) contingencies. Colored lines indicate means and error bars show ± 1 within-subject SEM. Black line shows chance accuracy in both plots. SEM = standard error of the mean. See the online article for the color version of this figure.

* $p < .05$. ** $p < .01$.

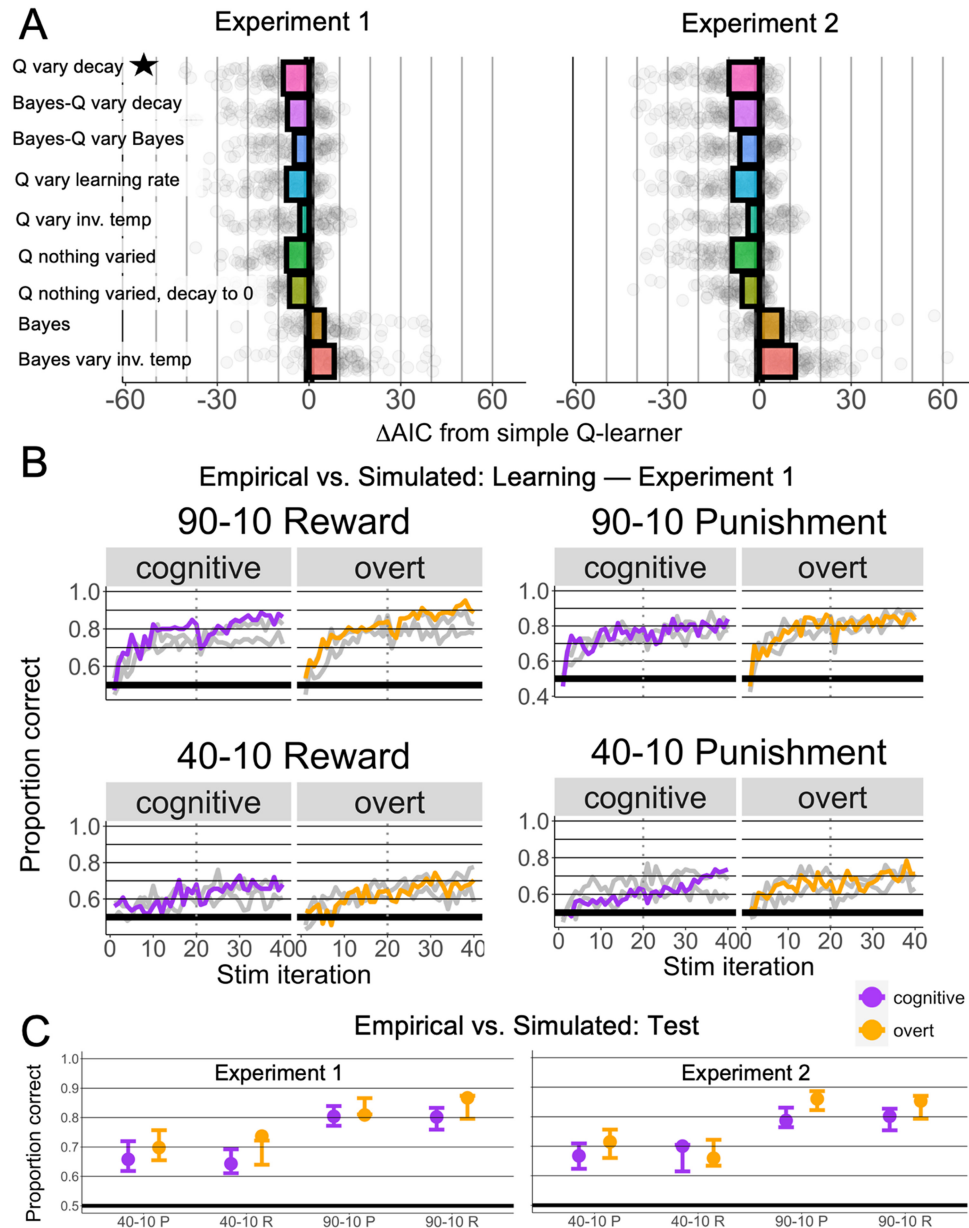
model comparison (i.e., fit to task data while penalizing for model complexity) and model validation (i.e., ability to capture key features of the data; “Method”). In both experiments, the most successful model (Figure 3) was a Q-learning RL model that learned values of the actions in each state (Q-values) through trial and error, but where the Q-values decayed (Brown et al., 2018; Collins & Frank, 2012; Collins et al., 2014; Hitchcock et al., 2022; Katahira &

Toyama, 2021; Niv et al., 2015) over trials toward a default prior via a decay parameter, ϕ ,

$$Q_{t+1} = (1 - \phi) Q_t + \phi Q_{\text{prior}}, \text{ for all } Q_t \text{ other than } Q(s, a)_t, \quad (1)$$

where $Q(s, a)_t$ was the value of the action selected on the current trial, Q_{prior} were uniform prior values that differed by valence (see

Figure 3
Computational Model Comparison and Validation



Note. (A) Model comparison plot showing the change in penalized model fit of nine plausible models of the task against a simple Q-learning model (the models were Bayesian, reinforcement-learning, and hybrid models; see the Supplemental Methods in the online supplemental materials for full details). Bar shows the average subject-level change in model fit and gray points show changes for each subject. The star indicates that the “Q vary decay” model provided the best fit to the data in both experiments (here shown as improvement relative to the simpler Q-learner model; see Table 1 in the online supplemental materials for total sum AIC of each model). (B) Model validation in the learning phase: empirical learning curves (colored lines) against simulated learning curves from 5th and 95th percentile accuracy (gray lines), thus showing the range of the simulated data, in all {condition-contingency} pairs. The learning curves shown are from Experiment 1; for the Experiment 2 plot, see Figure 3B in the online supplemental materials. (C) Model validation in the test phase: empirical means (points) and the range of 5th–95th percentile accuracy simulations (error bars). Black horizontal lines in B and C show chance accuracy. Q = Q-learner; inv. = inverse; AIC = Akaike information criterion; P = punishment; R = reward. See the online article for the color version of this figure.

model description in “Supplemental Methods” in the online supplemental materials for details), and where the ϕ parameter was allowed to vary by condition (ϕ^{cog} vs. ϕ^{overt}). This model that allowed ϕ to vary by condition provided a better fit to the data in both experiments than the same model with shared parameters across conditions (the “Q nothing varied” model in Figure 3A), which suggests that differences in the parameter help to account for differences between the conditions. It also improved fit compared to other Bayesian, RL, and hybrid models that we examined where other parameters were allowed to vary by condition (“Supplemental Methods” in the online supplemental materials). Further, across participants, the improvement in model fit relative to the “Q nothing varied” model correlated with the absolute difference in ϕ values between the conditions (Experiment 1 sum $\Delta\text{AIC} = -134.03$, Spearman’s $\rho = .67$, $p < 2e-16$; Experiment 2 sum $\Delta\text{AIC} = -147.22$, Spearman’s $\rho = .63$, $p < 2e-16$), which means that some participants showed a larger difference in decay between the conditions—and corresponding model fit improvement—than others. Full parameter values for the model are reported in Table 2 in the online supplemental materials.

Simulations from this model (“Method”) recapitulated various key features of the data, including learning curves and test phase accuracy across all eight {contingency-condition} combinations, decreases in accuracy as a function of delay in the learning phase, and the rate at which participants consistently picked the worst option in the test phase (Figure 3B and 3C; Figures 3 and 4 in the online supplemental materials).

A Difference in Retention of Action Values Helps to Account for Accuracy Differences Across Conditions

Having established that an RL model in which the ϕ parameter was allowed to vary by condition provided an adequate account of the data, we next evaluated whether this parameter differs between conditions and helps to explain individual differences in accuracy.

In both experiments, the ϕ parameter was higher (more rapid decay) in the cognitive than overt condition, with the majority of participants showing this pattern, although this difference was marginally significant in Experiment 1 (Figure 4A; Experiment 1: median $\phi^{\text{cog}} = .16$, median $\phi^{\text{overt}} = .13$; $\phi^{\text{cog}} > \phi^{\text{overt}} = 57.6\%$; $\chi^2 = 2.89$, $p = .089$; Experiment 2: $\phi^{\text{cog}} = .20$, median $\phi^{\text{overt}} = .16$; $\phi^{\text{cog}} > \phi^{\text{overt}} = 63.04\%$; $\chi^2 = 9.39$, $p < .003$). Interpretation of ϕ estimates and the differences therein are somewhat qualified by their having modest parameter recovery (Spearman’s ρ : $\phi^{\text{cog}} = .67$, $\phi^{\text{overt}} = .61$, $\phi^{\text{cog}} - \phi^{\text{overt}} = .56$; sign difference $\phi^{\text{cog}} - \phi^{\text{overt}} = 71.43\%$ correct; Figure 5 in the online supplemental materials). Thus, a target for future research will be to develop task designs where this parameter can be more precisely dissociated from other parameters. Nonetheless, in both experiments, participants’ differences in the condition-specific ϕ parameter estimates correlated with their accuracy differences between the conditions not only in the learning phase, but also in the test phase (Figure 4B). The correlations confirm that this parameter does not merely capture effects due to delay (or other differences) during initial learning, but also accuracy in the test phase by dictating the retention of action values that are drawn upon in this phase. Notably, the formation of robust expected actual values is thought to rely on working memory (Frank & Claus, 2006; Gold et al., 2012; Hernaus et al., 2018, 2019). Simulation using higher ϕ^{cog} than ϕ^{overt} confirmed that a difference in this parameter (with no other parameters varied between conditions) is sufficient to produce substantial accuracy

differences in the learning and test phases (Figure 6 in the online supplemental materials).

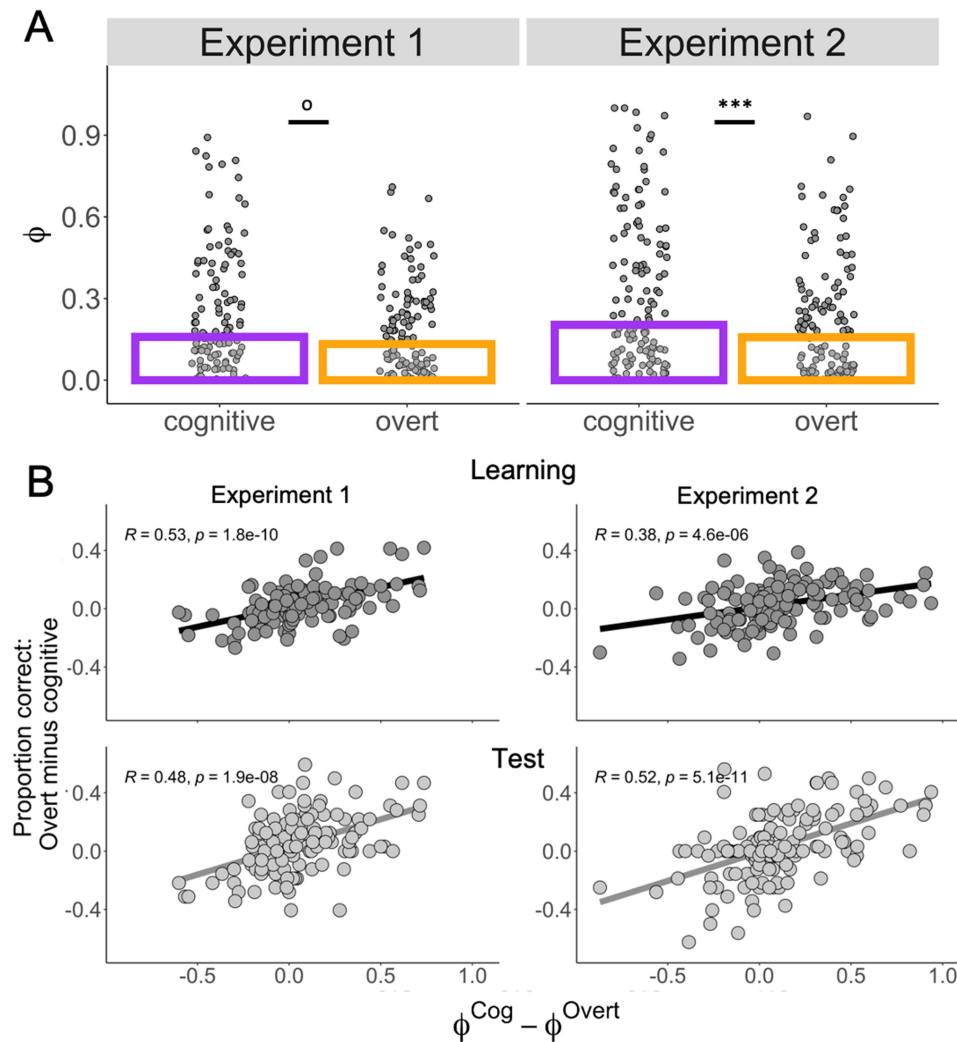
We did not find any significant relationships between measures of perseverative thinking (Perseverative Thinking Questionnaire and Rumination-Response Style questionnaire) and the decay rate in the cognitive condition or the difference in the decay rate between conditions ($ps > .07$). Thus, we found no evidence that people who especially struggled to learn cognitive action values were also especially prone to think in a repetitive negative way. However, note that this question is distinct from whether cognitive actions are in general more difficult to learn, as our findings do suggest. This difficulty could plausibly help to explain why it is difficult to acquire adaptive alternative cognitive actions—to replace currently ingrained ones—in psychotherapy; and why maladaptive thinking patterns are readily acquired by individuals with other individual differences and/or learning histories that confer risk (e.g., see Hitchcock & Frank, 2024). Please see the Supplemental Results in the online supplemental materials for further exploratory individual-differences analyses.

Discussion

Many prominent clinical-science theories assume that mental behaviors are similar to motor behaviors, in that they are maintained or extinguished based on their consequences. For instance, rumination is thought to be maintained by justifying abstention from one’s commitments (Nolen-Hoeksema et al., 2008), suicidal thinking and self-injury imagery by relieving negative affect (Coppersmith et al., 2023; Lawrence et al., 2023), and worry by replacing aversive mental imagery (Borkovec et al., 1993) or buffering against precipitous drops in affect (Newman & Llera, 2011). Many behavioral psychotherapies assume that, despite serving these functions, maladaptive mental behaviors are myopic or otherwise neglectful of negative consequences. Hence, these therapies strive to help people learn alternative mental behaviors with better long-run consequences than those in their current repertoire. Examples include thinking concretely rather than abstractly (Martell et al., 2021; Watkins, 2018); attending to the present moment instead of unproductively worrying (Orsillo & Roemer, 2007); “spoiling” instead of enacting mental compulsions (Foa et al., 2012); and recognizing thoughts as mere thoughts, which can be viewed from a mental distance and appraised for their utility, rather than acted upon as though they were imperatives (S. C. Hayes et al., 2011).

A key assumption in these approaches is that the consequences of mental behavior can be learned through experience—hence, mental behaviors with better consequences will eventually replace those with more negative ones. Yet, little is known about the foundational question of how learning adaptive mental behavior compares in difficulty to learning about motor behavior. We therefore designed a novel probabilistic learning task to compare the ability to learn optimal motor (overt) versus mental (cognitive) behaviors.

Our results neither suggest that both types of learning are comparably easy nor that learning adaptive cognitive (vs. overt) behavior is drastically more difficult for nearly everyone. Rather, across two experiments and in both learning and test phases, we found that, overall, people had more difficulty learning to take adaptive cognitive than overt actions. Yet, there was substantial heterogeneity among participants, with 33%–39% of participants actually showing higher accuracy in the cognitive than overt condition.

Figure 4*The Decay Parameter, ϕ* 

Note. (A) Parameter values in the cognitive and overt conditions in Experiments 1 and 2. Bars show median of all subjects', and points individual subjects', values for ϕ^{Cog} and ϕ^{Overt} . $o = p < .1$; $*** p < .005$. (B) Correlation between subjects' ϕ^{Cog} minus ϕ^{Overt} difference scores (positive values correspond to relatively higher decay in the cognitive condition) and differences in their overt minus cognitive accuracy. Positive correlations indicate that those with relatively higher decay in the cognitive (than overt) condition tended to have higher accuracy in the overt (than cognitive) condition, in both the learning and test phases. Cog = cognitive. See the online article for the color version of this figure.

To understand the overall accuracy difference between conditions, we built various computational models. The most successful model, in terms of its (penalized) fit to the data and ability to capture key behavioral patterns, was an RL model that learned action values that subsequently decayed across trials—with the rate of decay varied by condition. We found that between-condition differences in the decay rate were closely correlated with between-condition differences in accuracy—in both the learning and the test phase (even though decay only occurred during the learning phase). Moreover, decay was elevated in the cognitive (relative to overt) condition in both experiments, although the difference was not statistically significant in Experiment 1. Elevated decay in the model suggests more difficulty

with credit assignment because it means less robust action values tended to be available on a given trial during the learning phase and ultimately during the test phase. Impaired credit assignment may have occurred in the cognitive condition because constructing expected values relies on working memory (Frank & Claus, 2006; Geana et al., 2022; Gold et al., 2012; Hernaus et al., 2018), which is also required to effect a cognitive action. A complementary possibility is that the interposition of a cognitive operation between action initiation and reinforcement, which definitionally occurs when taking a cognitive action, interferes with credit assignment.

Crucially, the pattern of our behavioral findings did not suggest that participants in the cognitive condition simply failed to hold

contingencies in working memory across trials during learning, but rather suggested impairment in constructing long-run expected values for actions—a process also thought to rely on working memory (Frank & Claus, 2006; Gold et al., 2012; Hernaus et al., 2018, 2019). That is because we found weaker performance in the cognitive condition not only in the learning phase but also in a test phase, at which point working-memory representations that merely temporarily represented contingencies over short delays should have long since dissipated.

A key strength of our experiments is that they address questions of crucial practical importance in clinical science and psychotherapy: whether and why learning about cognitive actions may be especially difficult. We conducted two experiments with different task variants and found consistent behavioral and computational modeling results. Our computational model was able to capture numerous key features of the data, and it offered insight into why accuracy differed (on average) between conditions by pointing to disruption of the ability to form and retain expected values. Although this is only a first investigation that raises many mechanistic questions (described below), this line of work might eventually offer ideas for “rescuing” credit assignment to accelerate the acquisition of adaptive mental behaviors in psychotherapy. For instance, it may be beneficial to shape these behaviors (see Krueger & Dayan, 2009; see also Hitchcock & Frank, 2024), acquire them in settings that minimize distraction and maximize the salience of feedback, or promote explicit recollection of psychotherapy content that might serve as cues to practice alternative mental behaviors repeatedly in everyday life (Harvey et al., 2014). The heterogeneity between participants also opens the possibility of tailoring treatments that rely on learning adaptive cognitive actions. For instance, people who struggle with such learning may benefit from special assistance and scaffolding, whereas people adept at such learning may be especially successful in treatments that capitalize on that skill (see Cohen & DeRubeis, 2018, p. 115).

However, this first investigation also raises a number of questions about the granular mechanisms that make it difficult to form expected values for cognitive actions. Although we postulated possibilities including dual demands on working memory and incorrect credit ascription, other (possibly complementary) possibilities include the added time necessary to effect a cognitive action and that cognitive actions do not afford sensory cues (which might aid in representing an action as such and thereby facilitate credit assignment). In future work, we plan to directly manipulate various factors that may make cognitive-action learning more difficult to pinpoint the responsible mechanism(s) in more detail.

Our study bears most directly on clinical theories that assume mental behavior serve a function (i.e., are responsive to reinforcement; e.g., Borkovec et al., 1993; Coppersmith et al., 2023; Lawrence et al., 2023; Newman & Llera, 2011; Nolen-Hoeksema et al., 2008) and to psychotherapies that take a behavioral approach to cognitive activity (e.g., Martell et al., 2021; Orsillo & Roemer, 2007; S. C. Hayes et al., 2011; Watkins, 2018). These psychotherapies also assume that mental behaviors serve a function, and engineer strategies to decrease the frequency of maladaptive mental behaviors (such as rumination) or increase the frequency of adaptive mental behaviors (such as mindfully attending to the present moment). Our work connects these clinical theories and therapies to RL research concerning how adaptive cognitive actions (such as gating items into and out of working memory) are learned and executed, irrespective of the item’s specific content (e.g., Braver et al., 1999; Chatham & Badre, 2015; Dayan, 2012; Frank & Badre, 2012; O’Reilly et al., 1999; O’Reilly & Frank, 2006; Rac-Lubashevsky &

Frank, 2021; Todd et al., 2009; Trutti et al., 2021; Westbrook & Braver, 2016; see also, Hitchcock & Frank, 2024). By contrast, our approach is not as naturally compatible with cognitive-behavioral therapies that focus on changing cognitive content, whether directly (e.g., through cognitive restructuring of beliefs) or indirectly by modifying overt behavior in the service of cognitive change (e.g., exposure therapy designed to change beliefs about safety). How to reconcile the influence of cognitive content (e.g., explicit beliefs) on overt behavior and vice versa, with research on how adaptive cognitive actions themselves are acquired via reinforcement, is an open question for future research (see Atlas et al., 2016; Berwian et al., 2023; Dercon et al., 2024; Doll et al., 2009 regarding the relationship between rules or beliefs and RL).

A limitation of our experiments is that we did not include measures of working-memory capacity that might co-vary with impairment in the cognitive condition. Such findings would have lent credence to the possibility that dual working-memory demands are responsible for the impairment. More generally, we were unable here to parse the considerable heterogeneity between subjects (e.g., finding individual differences), limiting the conclusions that we can currently draw about how much of the between-condition effect variability was meaningful versus simply because of noisy aspects of the task. Another caveat is that we deliberately designed the cognitive actions in our task to involve maintaining and manipulating mental content. Many real-world cognitive actions share this demand, but we do not expect our results to generalize to mental processes (e.g., orienting attention) that do not.

An important nuance is that, when psychotherapies seek to replace maladaptive mental behaviors with more adaptive ones, they may often need to contend with deeply ingrained mental behaviors that have fallen out of goal-directed control—i.e., become habitual (Brewer, 2021; Watkins & Nolen-Hoeksema, 2014). How the challenge of acquiring adaptive mental behavior documented here influences the habitization of mental behavior, and the ability to break mental habits, are important topics for future research. Moreover, although we employed a number of computational modeling best practices (including extensive model validation, model comparison, and parameter recovery) and regularized subject-level estimates via group-level information using a hierarchical modeling procedure that can improve subject-level estimates (Katahira, 2016), a direction for future work is to employ hierarchical Bayesian modeling and new task designs that might further improve estimation and parameter recovery.

Finally, we wish to emphasize that this study examined the comparative challenge of adaptive learning of quite simple mental versus motor behaviors—which differed only in that the former required performing a cognitive operation (given that carrying out a mental behavior by definition entails performing one or more cognitive operations). Our design followed the logic of many past RL studies, which assume that it is possible to gain insights into the mechanics of learning in highly simplified but tightly controlled experimental designs that generalize to complex behaviors in everyday life (reviewed in Niv, 2009; Rmus et al., 2021; and see Frey et al., 2021; Kasanova et al., 2017 for examples of such tasks predicting real-life behavior). Given that our study was, to our knowledge, the first to directly compare the ability to learn adaptive mental versus motor behaviors, we matched these conditions as closely as possible. In particular, we deliberately avoided adding emotional content, or multiple operations, to the cognitive condition. We made this choice although maladaptive thinking patterns, as well as many skills taught in psychotherapy such as cognitive

restructuring, are clearly multistage valenced processes. As such, the initial evidence presented here—that mental (vs. motor) behaviors may be more intrinsically difficult to learn, even in this elemental form—is undoubtedly just one of a number of reasons that people are apt to acquire maladaptive thinking patterns and why it is difficult to ingrain replacement thinking patterns within psychotherapy.

Notably, we recently developed a complementary theoretical account of rumination and worry in which the challenge of learning adaptive mental behaviors was just one part of the theory (Hitchcock & Frank, 2024). In particular, drawing on computational models that learn adaptive working-memory operations in sequential tasks, we proposed that rumination and worry arise from an attempt at problem solving gone awry at one or more of four distinct stages: first, selecting an overarching hypothesis (that might instigate a chain of rumination if it is open-ended and/or negatively valenced, e.g., “I am socially incompetent”); second, executing subproblems related to it (e.g., thinking of specific instances of social incompetence with friends); third, switching between subproblems; and fourth, reinforcing subproblems and/or the overarching hypothesis. This theory casts rumination and worry as arising from distinct but interacting challenges involved in selecting, executing, switching between, and learning the consequences of a hierarchical, sequential set of mental operations. The possible intrinsic difficulty of adaptively learning (even a single) mental operation, suggested by our results here, implies a potentially inherent difficulty as the reinforcement stage. It may also suggest therapeutic techniques for simplifying the acquisition of adaptive alternative behaviors, such as shaping and scaffolding (Hitchcock & Frank, 2024, p. 4). Yet, according to our theory (Hitchcock & Frank, 2024), various other factors and individual differences likely contribute to the tendency to ruminate or worry (e.g., abstract thinking, stable mental content, negatively valenced thinking; van Vugt et al., 2012; Watkins, 2008; Whitmer & Gotlib, 2013). Likewise, complex multistage processes taught in psychotherapy (such as cognitive restructuring) presumably rely on various interacting factors and individual differences (Hitchcock & Frank, 2024, p. 4). In short, our work here—that suggests a credit assignment challenge even when learning elemental mental (vs. motor) behaviors—helps to lay a foundation for broader theories and task variations necessary for more complete accounts of various clinically relevant multistage, valenced thinking patterns.

In sum, in a novel task, we found that people had more difficulty learning to select optimal cognitive (vs. overt) actions, with computational modeling tracing this difficulty to impaired formation and retention of expected action values. These findings pave the way for future research into individual differences, more granular mechanisms of this impairment, and multistage and valenced mental behaviors—with numerous potential applications, including to clinical theories and psychotherapies that aim to help people learn to think in healthier and more productive ways.

References

- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199–213). Springer. https://doi.org/10.1007/978-1-4612-1694-0_15
- Asaad, W. F., Lauro, P. M., Perge, J. A., & Eskandar, E. N. (2017). Prefrontal neurons encode a solution to the credit-assignment problem. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 37(29), 6995–7007. <https://doi.org/10.1523/JNEUROSCI.3311-16.2017>
- Atlas, L. Y., Doll, B. B., Li, J., Daw, N. D., & Phelps, E. A. (2016). Instructed knowledge shapes feedback-driven aversive learning in striatum and orbitofrontal cortex, but not the amygdala. *eLife*, 5, Article e15192. <https://doi.org/10.7554/eLife.15192>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014, June 23). *Fitting linear mixed-effects models using lme4*. arXiv. <https://doi.org/10.48550/arXiv.1406.5823>
- Berwian, I., Hitchcock, P., Pisupati, S., Schoen, G., & Niv, Y. (2023, August 24). *Using learning theories to advance psychotherapy theory and research*. <https://doi.org/10.31234/osf.io/8snbq>
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.
- Borkovec, T. D., Lyonfields, J. D., Wiser, S. L., & Deihl, L. (1993). The role of worrisome thinking in the suppression of cardiovascular response to phobic imagery. *Behaviour Research and Therapy*, 31(3), 321–324. [https://doi.org/10.1016/0005-7967\(93\)90031-o](https://doi.org/10.1016/0005-7967(93)90031-o)
- Braver, T. S., Barch, D. M., & Cohen, J. D. (1999). Cognition and control in schizophrenia: A computational model of dopamine and prefrontal function. *Biological Psychiatry*, 46(3), 312–328. [https://doi.org/10.1016/s0006-3223\(99\)00116-x](https://doi.org/10.1016/s0006-3223(99)00116-x)
- Brewer, J. (2021). *Unwinding anxiety: New science shows how to break the cycles of worry and fear to heal your mind*. Penguin Random House LLC.
- Brown, V. M., Zhu, L., Wang, J. M., Frueh, B. C., King-Casas, B., & Chiu, P. H. (2018). Associability-modulated loss learning is increased in posttraumatic stress disorder. *eLife*, 7, Article e30150. <https://doi.org/10.7554/eLife.30150>
- Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, 67(2), 319–333. <https://doi.org/10.1037/0022-3514.67.2.319>
- Casella, G. (1985). An introduction to empirical Bayes data analysis. *The American Statistician*, 39(2), 83–87. <https://doi.org/10.1080/00031305.1985.10479400>
- Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. *Current Opinion in Behavioral Sciences*, 1, 23–31. <https://doi.org/10.1016/j.cobeha.2014.08.001>
- Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment selection in depression. *Annual Review of Clinical Psychology*, 14(1), 209–236. <https://doi.org/10.1146/annurev-clinpsy-050817-084746>
- Collins, A. G. E., Albrecht, M. A., Waltz, J. A., Gold, J. M., & Frank, M. J. (2017). Interactions among working memory, reinforcement learning, and effort in value-based choice: A new paradigm and selective deficits in schizophrenia. *Biological Psychiatry*, 82(6), 431–439. <https://doi.org/10.1016/j.biopsych.2017.05.017>
- Collins, A. G. E., Brown, J. K., Gold, J. M., Waltz, J. A., & Frank, M. J. (2014). Working memory contributions to reinforcement learning impairments in schizophrenia. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 34(41), 13747–13756. <https://doi.org/10.1523/JNEUROSCI.0989-14.2014>
- Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *The European Journal of Neuroscience*, 35(7), 1024–1035. <https://doi.org/10.1111/j.1460-9568.2011.07980.x>
- Collins, A. G. E., & Frank, M. J. (2018). Within-and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. *Proceedings of the National Academy of Sciences*, 115(10), 2502–2507. <https://doi.org/10.1073/pnas.1720963115>
- Coppersmith, D. D. L., Millgram, Y., Kleiman, E. M., Fortgang, R. G., Millner, A. J., Frumkin, M. R., Bentley, K. H., & Nock, M. K. (2023). Suicidal thinking as affect regulation. *Journal of Psychopathology and Clinical Science*, 132(4), 385–395. <https://doi.org/10.1037/abn0000828>

- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6), 1204–1215. <https://doi.org/10.1016/j.neuron.2011.02.027>
- Dayan, P. (2012). How to set the switches on this thing. *Current Opinion in Neurobiology*, 22(6), 1068–1074. <https://doi.org/10.1016/j.conb.2012.05.011>
- Dercon, Q., Mehrhof, S. Z., Sandhu, T. R., Hitchcock, C., Lawson, R. P., Pizzagalli, D. A., Dalgleish, T., & Nord, C. L. (2024). A core component of psychological therapy causes adaptive changes in computational learning mechanisms. *Psychological Medicine*, 54(2), 327–337. <https://doi.org/10.1017/S0033291723001587>
- Doll, B. B., Jacobs, W. J., Sanfey, A. G., & Frank, M. J. (2009). Instructional control of reinforcement learning: A behavioral and neurocomputational investigation. *Brain Research*, 1299, 74–94. <https://doi.org/10.1016/j.brainres.2009.07.007>
- Ehring, T., Zetsche, U., Weidacker, K., Wahl, K., Schönfeld, S., & Ehlers, A. (2011). The Perseverative Thinking Questionnaire (PTQ): Validation of a content-independent measure of repetitive negative thinking. *Journal of Behavior Therapy and Experimental Psychiatry*, 42(2), 225–232. <https://doi.org/10.1016/j.jbtep.2010.12.003>
- Foa, E. B., Yadin, E., & Lichner, T. K. (2012). *Exposure and response (ritual) prevention for obsessive-compulsive disorder: Therapist guide*. Oxford University Press.
- Frank, M. J., & Badre, D. (2012). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: Computational analysis. *Cerebral Cortex*, 22(3), 509–526. <https://doi.org/10.1093/cercor/bhr114>
- Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological Review*, 113(2), 300–326. <https://doi.org/10.1037/0033-295X.113.2.300>
- Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science*, 306(5703), 1940–1943. <https://doi.org/10.1126/science.1102941>
- Frey, A.-L., Frank, M. J., & McCabe, C. (2021). Social reinforcement learning as a predictor of real-life experiences in individuals with high and low depressive symptomatology. *Psychological Medicine*, 51(3), 408–415. <https://doi.org/10.1017/S0033291719003222>
- Geana, A., Barch, D. M., Gold, J. M., Carter, C. S., MacDonald, A. W., Ragland, J. D., Silverstein, S. M., & Frank, M. J. (2022). Using computational modeling to capture schizophrenia-specific reinforcement learning differences and their implications on patient classification. *Biological psychiatry. Cognitive Neuroscience and Neuroimaging*, 7(10), 1035–1046. <https://doi.org/10.1016/j.bpsc.2021.03.017>
- Ghalanos, A., & Theussl, S. (2011). *Package "Rsolnp."* <https://civil.colorado.edu/~balajir/CVEN5393/R-sessions/sess1/Rsolnp-1.pdf>
- Gold, J. M., Waltz, J. A., Matveeva, T. M., Kasanova, Z., Strauss, G. P., Herbener, E. S., Collins, A. G. E., & Frank, M. J. (2012). Negative symptoms and the failure to represent the expected reward value of actions: Behavioral and computational modeling evidence. *Archives of General Psychiatry*, 69(2), 129–138. <https://doi.org/10.1001/archgenpsychiatry.2011.1269>
- Guitart-Masip, M., Huys, Q. J. M., Fuentemilla, L., Dayan, P., Duzel, E., & Dolan, R. J. (2012). Go and no-go learning in reward and punishment: Interactions between affect and effect. *NeuroImage*, 62(1), 154–166. <https://doi.org/10.1016/j.neuroimage.2012.04.024>
- Hamid, A. A., Frank, M. J., & Moore, C. I. (2021). Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell*, 184(10), 2733–2749.e16. <https://doi.org/10.1016/j.cell.2021.03.046>
- Harvey, A. G., Lee, J., Williams, J., Hollon, S. D., Walker, M. P., Thompson, M. A., & Smith, R. (2014). Improving outcome of psychosocial treatments by enhancing memory and learning. *Perspectives on Psychological Science*, 9(2), 161–179. <https://doi.org/10.1177/1745691614521781>
- Hayes, S. A., Orsillo, S. M., & Roemer, L. (2010). Changes in proposed mechanisms of action during an acceptance-based behavior therapy for generalized anxiety disorder. *Behaviour Research and Therapy*, 48(3), 238–245. <https://doi.org/10.1016/j.brat.2009.11.006>
- Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2011). *Acceptance and commitment therapy, second edition: The process and practice of mindful change*. Guilford Press.
- Hernaus, D., Frank, M. J., Brown, E. C., Brown, J. K., Gold, J. M., & Waltz, J. A. (2019). Impaired expected value computations in schizophrenia are associated with a reduced ability to integrate reward probability and magnitude of recent outcomes. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4(3), 280–290. <https://doi.org/10.1016/j.bpsc.2018.11.011>
- Hernaus, D., Gold, J. M., Waltz, J. A., & Frank, M. J. (2018). Impaired expected value computations coupled with overreliance on stimulus-response learning in schizophrenia. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, 3(11), 916–926. <https://doi.org/10.1016/j.bpsc.2018.03.014>
- Hitchcock, P., Forman, E., Rothstein, N., Zhang, F., Kounios, J., Niv, Y., & Sims, C. (2022). Rumination derails reinforcement learning with possible implications for ineffective behavior. *Clinical Psychological Science*, 10(4), 714–733. <https://doi.org/10.1177/21677026211051324>
- Hitchcock, P., & Frank, M. (2024). From tripping and falling to ruminating and worrying: A meta-control account of repetitive negative thinking. *Current Opinion in Behavioral Sciences*, 56, Article 101356. <https://doi.org/10.1016/j.cobeha.2024.101356>
- Jocham, G., Brodersen, K. H., Constantinescu, A. O., Kahn, M. C., Ianni, A. M., Walton, M. E., Rushworth, M. F., & Behrens, T. E. (2016). Reward-guided learning with and without causal attribution. *Neuron*, 90(1), 177–190. <https://doi.org/10.1016/j.neuron.2016.02.018>
- Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. <https://doi.org/10.1126/science.1152408>
- Kasanova, Z., Ceccarini, J., Frank, M. J., van Amelsvoort, T., Booi, J., Heinzl, A., Mottaghy, F., & Myin-Germeys, I. (2017). Striatal dopaminergic modulation of reinforcement learning predicts reward-oriented behavior in daily life. *Biological Psychology*, 127, 1–9. <https://doi.org/10.1016/j.biopsycho.2017.04.014>
- Katahira, K. (2016). How hierarchical models improve point estimates of model parameters at the individual level. *Journal of Mathematical Psychology*, 73, 37–58. <https://doi.org/10.1016/j.jmp.2016.03.007>
- Katahira, K., & Toyama, A. (2021). Revisiting the importance of model fitting for model-based fMRI: It does matter in computational psychiatry. *PLoS Computational Biology*, 17(2), Article e1008738. <https://doi.org/10.1371/journal.pcbi.1008738>
- Kazdin, A. E. (2012). *Behavior modification in applied settings*. Waveland Press.
- Krueger, K. A., & Dayan, P. (2009). Flexible shaping: How learning in small steps helps. *Cognition*, 110(3), 380–394. <https://doi.org/10.1016/j.cognition.2008.11.014>
- Lamba, A., Nassar, M. R., & FeldmanHall, O. (2023). Prefrontal cortex state representations shape human credit assignment. *Elife*, 12, Article e84888. <https://doi.org/10.7554/eLife.84888>
- Lawrence, H. R., Balkind, E. G., Ji, J. L., Burke, T. A., & Liu, R. T. (2023). Mental imagery of suicide and non-suicidal self-injury: A meta-analysis and systematic review. *Clinical Psychology Review*, 103, Article 102302. <https://doi.org/10.1016/j.cpr.2023.102302>
- Martell, C. R., Dimidjian, S., & Herman-Dunn, R. (2021). *Behavioral activation for depression, second edition: A clinician's guide*. Guilford Publications.
- Newman, M. G., & Llera, S. J. (2011). A novel theory of experiential avoidance in generalized anxiety disorder: A review and synthesis of research supporting a contrast avoidance model of worry. *Clinical Psychology Review*, 31(3), 371–382. <https://doi.org/10.1016/j.cpr.2011.01.008>
- Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3), 139–154. <https://doi.org/10.1016/j.jmp.2008.12.005>

- Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *Journal of Neuroscience*, 35(21), 8145–8157. <https://doi.org/10.1523/JNEUROSCI.2978-14.2015>
- Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 3(5), 400–424. <https://doi.org/10.1111/j.1745-6924.2008.00088.x>
- O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 375–411). Cambridge University Press.
- 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. <https://doi.org/10.1162/089976606775093909>
- Orsillo, S. M., & Roemer, L. (2007). *Acceptance- and mindfulness-based approaches to anxiety: Conceptualization and treatment*. Springer Science & Business Media.
- Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, 442(7106), 1042–1045. <https://doi.org/10.1038/nature05051>
- Piray, P., Dezfouli, A., Heskes, T., Frank, M. J., & Daw, N. D. (2019). Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. *PLoS Computational Biology*, 15(6), Article e1007043. <https://doi.org/10.1371/journal.pcbi.1007043>
- Provenza, N. R., Gelin, L. F. F., Mahaphanit, W., McGrath, M. C., Dastin-van Rijn, E. M., Fan, Y., Dhar, R., Frank, M. J., Restrepo, M. I., Goodman, W. K., & Borton, D. A. (2022). Honeycomb: A template for reproducible psychophysiological tasks for clinic, laboratory, and home use. *Brazilian Journal of Psychiatry*, 44(2), 147–155. <https://doi.org/10.1590/1516-4446-2020-1675>
- Rac-Lubashevsky, R., Cremer, A., Collins, A. G., Frank, M. J., & Schwabe, L. (2023). Neural index of reinforcement learning predicts improved stimulus–response retention under high working memory load. *Journal of Neuroscience*, 43(17), 3131–3143. <https://doi.org/10.1523/JNEUROSCI.1274-22.2023>
- Rac-Lubashevsky, R., & Frank, M. J. (2021). Analogous computations in working memory input, output and motor gating: Electrophysiological and computational modeling evidence. *PLoS Computational Biology*, 17(6), Article e1008971. <https://doi.org/10.1371/journal.pcbi.1008971>
- Radulescu, A., Daniel, R., & Niv, Y. (2016). The effects of aging on the interaction between reinforcement learning and attention. *Psychology and Aging*, 31(7), 747–757. <https://doi.org/10.1037/pag0000112>
- Ramnero, J., & Tömeke, N. (2008). *The ABCs of human behavior: Behavioral principles for the practicing clinician*. New Harbinger Publications.
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rmus, M., McDougle, S. D., & Collins, A. G. (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>
- Schoenbaum, G., Setlow, B., Nugent, S. L., Saddoris, M. P., & Gallagher, M. (2003). Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals. *Learning & Memory*, 10(2), 129–140. <https://doi.org/10.1101/lm.55203>
- Shahar, N., Moran, R., Hauser, T. U., Kievit, R. A., McNamee, D., Moutoussis, M., Dolan, R. J., Bullmore, E., Dolan, R., Goodyer, I., Fonagy, P., Jones, P., Moutoussis, M., Hauser, T., Neufeld, S., Romero-Garcia, R., St Clair, M., Vértes, P., Whitaker, K., ... the NSPN Consortium. (2019). Credit assignment to state-independent task representations and its relationship with model-based decision making. *Proceedings of the National Academy of Sciences*, 116(32), 15871–15876. <https://doi.org/10.1073/pnas.1821647116>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning, second edition: An introduction*. MIT Press. <https://play.google.com/store/books/details?id=uWV0DwAAQBAJ>
- Todd, M. T., Niv, Y., & Cohen, J. D. (2009). Learning to use working memory in partially observable environments through dopaminergic reinforcement. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems* (Vol. 21, pp. 1689–1696). Cambridge.
- Treynor, W., Gonzalez, R., & Nolen-Hoeksema, S. (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research*, 27(3), 247–259. <https://doi.org/10.1023/A:1023910315561>
- Trutti, A. C., Verschooren, S., Forstmann, B. U., & Boag, R. J. (2021). Understanding subprocesses of working memory through the lens of model-based cognitive neuroscience. *Current Opinion in Behavioral Sciences*, 38, 57–65. <https://doi.org/10.1016/j.cobeha.2020.10.002>
- Van Vugt, M. K., Hitchcock, P., Shahar, B., & Britton, W. (2012). The effects of mindfulness-based cognitive therapy on affective memory recall dynamics in depression: A mechanistic model of rumination. *Frontiers in Human Neuroscience*, 6, Article 257. <https://doi.org/10.3389/fnhum.2012.00257>
- Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin*, 134(2), 163–206. <https://doi.org/10.1037/0033-2909.134.2.163>
- Watkins, E. R. (2018). *Rumination-focused cognitive-behavioral therapy for depression*. Guilford Press.
- Watkins, E. R., & Nolen-Hoeksema, S. (2014). A habit-goal framework of depressive rumination. *Journal of Abnormal Psychology*, 123(1), 24–34. <https://doi.org/10.1037/a0035540>
- Westbrook, A., & Braver, T. S. (2016). Dopamine does double duty in motivating cognitive effort. *Neuron*, 91(3), Article 708. <https://doi.org/10.1016/j.neuron.2016.07.020>
- Whitmer, A. J., & Gotlib, I. H. (2013). An attentional scope model of rumination. *Psychological Bulletin*, 139(5), 1036–1061. <https://doi.org/10.1037/a0030923>

Received September 26, 2023

Revision received April 9, 2024

Accepted April 13, 2024 ■