Biological Psychiatry

Interactions Among Working Memory, Reinforcement Learning, and Effort in Value-Based Choice: A New Paradigm and Selective Deficits in Schizophrenia

Anne G.E. Collins, Matthew A. Albrecht, James A. Waltz, James M. Gold, and Michael J. Frank

ABSTRACT

BACKGROUND: When studying learning, researchers directly observe only the participants' choices, which are often assumed to arise from a unitary learning process. However, a number of separable systems, such as working memory (WM) and reinforcement learning (RL), contribute simultaneously to human learning. Identifying each system's contributions is essential for mapping the neural substrates contributing in parallel to behavior; computational modeling can help to design tasks that allow such a separable identification of processes and infer their contributions in individuals.

METHODS: We present a new experimental protocol that separately identifies the contributions of RL and WM to learning, is sensitive to parametric variations in both, and allows us to investigate whether the processes interact. In experiments 1 and 2, we tested this protocol with healthy young adults (n = 29 and n = 52, respectively). In experiment 3, we used it to investigate learning deficits in medicated individuals with schizophrenia (n = 49 patients, n = 32 control subjects).

RESULTS: Experiments 1 and 2 established WM and RL contributions to learning, as evidenced by parametric modulations of choice by load and delay and reward history, respectively. They also showed interactions between WM and RL, where RL was enhanced under high WM load. Moreover, we observed a cost of mental effort when controlling for reinforcement history: participants preferred stimuli they encountered under low WM load. Experiment 3 revealed selective deficits in WM contributions and preserved RL value learning in individuals with schizophrenia compared with control subjects.

CONCLUSIONS: Computational approaches allow us to disentangle contributions of multiple systems to learning and, consequently, to further our understanding of psychiatric diseases.

Keywords: Computational modeling, Decision making, Effort, Reinforcement learning, Schizophrenia, Working memory

http://dx.doi.org/10.1016/j.biopsych.2017.05.017

Multiple neurocognitive systems interact to support various forms of learning, each with its own strengths and limitations. As experimenters, we can only observe the net behavioral outcome of the multifaceted learning process; thus, understanding how different systems contribute to learning in parallel requires creating experimental designs that can disentangle their contributions over different learning conditions. Some research has focused on the separable contributions of goal-directed planning versus stimulus–response habit formation during sequential multistage reinforcement learning (RL) (1–6). However, these processes can interact and, moreover, can themselves be subdivided into multiple systems; for example, planning involves working memory (WM), accurate representation of environmental contingencies, guided strategic search through such contingencies to determine a desired course of action, and so on.

We have previously shown that, even in simple stimulus–action–outcome learning situations with minimal demands on planning and search, there are dissociable contributing processes of WM and RL (7,8). We refer to working memory as a system that actively maintains information (such as the correct action to take in response to a given stimulus) in the face of interference (multiple intervening trials). WM is characterized by the limited availability of this information, due to either capacity or resource limitation, and decay/forgetting (9–12). We refer to reinforcement learning as the process that uses reward prediction errors (RPEs) to incrementally learn stimulus–response reward values in order to optimize expected future reward (13). These two systems have largely been studied in isolation, with WM depending on parietal/prefrontal cortex function (14–16) and RL relying on phasic dopaminergic signals conveying RPEs that modulate

ISSN: 0006-3223

corticostriatal synaptic plasticity (17,18). However, how both systems jointly contribute to learning, and whether and how they interact during learning, is currently poorly understood.

We developed an experimental protocol to highlight the role of WM in tasks typically considered to be under the purview of model-free RL (7). Specifically, we showed that learning from reward was affected by set size (the number of stimulus items presented during a block of trials) and delay (the number of intervening trials before a participant had a chance to reuse information). The effects of both load and delay decreased with repeated presentations, indicating a potential shift from early reliance on the faster but capacity-limited WM to later dominance of the RL system when associations became habituated. Our previous work showed that parsing out the components of learning can identify selective individual differences in healthy young adults (7) or deficits in clinical populations (8). However, it remained possible in this work that the paradigm was simply more parametrically sensitive to demands of WM and comparatively insensitive to the signature demands of RL. That is, in the deterministic environment, there was no need to learn precise estimates of reward probabilities for stimuli or actions.

Here, we present an improved learning task with more comparable sensitivity across WM and RL systems, providing firmer ground for their quantitative assessment. The design of the current task (Figure 1A–C) was motivated by prior modeling of WM and RL contributions to learning (Figure 1D, E) and extends our previous design with two new features. First, we added probabilistic variation in reward magnitudes (1 point vs. 2 points) across stimuli (Figure 1 A, B). Second, we added a subsequent surprise test phase (Figure 1C), affording the opportunity to assess whether choices were sensitive to parametric differences in values learned by RL [e.g., (19–21)]. We anticipated that the combination of these new features would allow us to investigate RL-based contributions to learning more directly in addition to the contribution of WM (Figure 1D). Furthermore, this improved task allows us to investigate whether WM demands during learning also influence the degree of value learning (Figure 1E). Such interactions would motivate refinement of existing computational models, which assume that RL and WM processes proceed independently and compete only for behavioral output (1,7).

To exemplify the utility of this new task in computational psychiatry research, we administered our new paradigm to people with schizophrenia (PSZ) and healthy control subjects (HCs) matched on important demographic variables (Table 1). The literature remains unclear as to the specific nature of learning impairments in PSZ (22). Indeed, recent studies

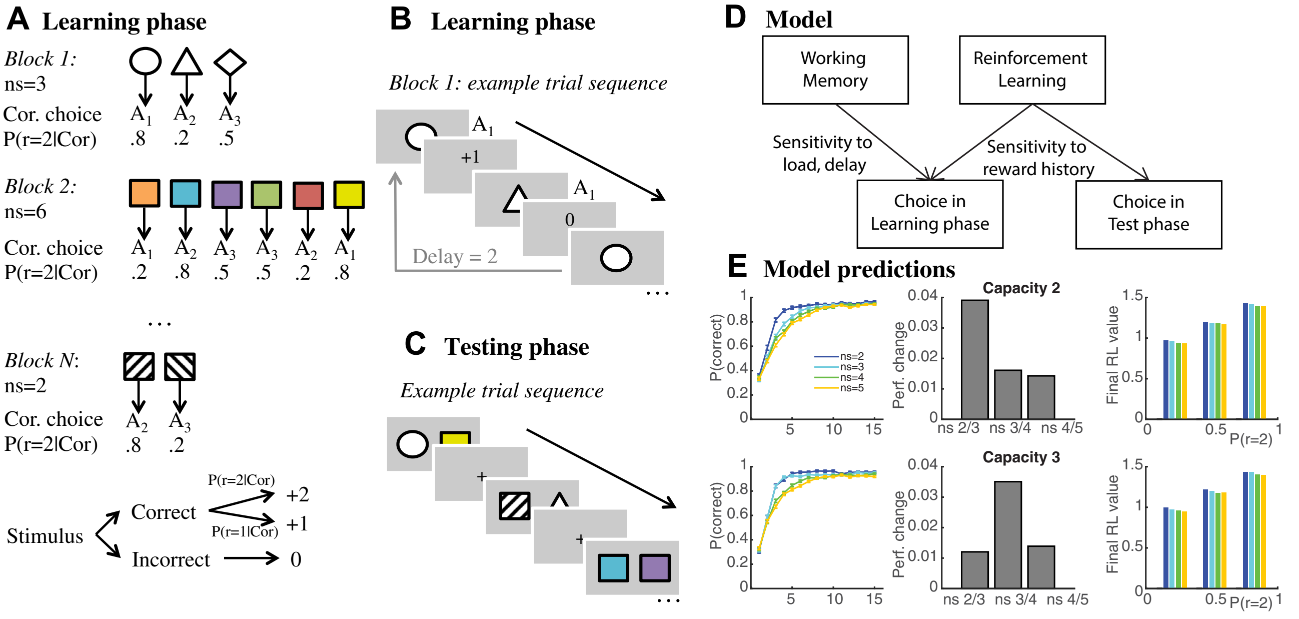

Figure 1. Experimental protocol. **(A)** Learning phase. Participants learn to select one of three actions (key presses $A_{1=3}$) for each stimulus in a block using reward feedback. Incorrect choices lead to feedback 0, while correct choices lead to reward, either +1 or +2 points, probabilistically. The probability of obtaining 2 points vs. 1 point is fixed for each stimulus, drawn from the set of (0.2, 0.5, or 0.8). The number of stimuli in a block (set size *ns*) varies from 1 to 6. **(B)** In learning blocks, stimuli are presented individually, randomly intermixed. Delay indicates the number of trials that occurred since the last correct choice for the current stimulus. **(C)** During a surprise test phase following learning, participants are asked to choose the more rewarding stimulus among pairs of previously encountered stimuli without feedback. **(D)** The computational model assumes that choice during learning comes from two separate systems, working memory (WM) and reinforcement learning (RL), making behavior sensitive to load, delay, and reward history. In contrast, test performance is dependent only on RL, such that if RL and WM are independent, choice should depend only on reward history. **(E)** A total of 100 simulations of the computational model with the new design for two sets of parameters representing poor WM use (capacity 2) and good WM use (capacity 3), respectively. (Left panel) Learning curves indicate the proportion of correct trials as a function of the number of encounters with given stimuli in different set sizes. (Middle panel) Difference in overall proportion of correct choices between subsequent set sizes shows a maximal drop in performance between set sizes 2 and 3 with capacity 2, while the drop is maximal between set sizes 3 and 4 with capacity 3. (Right panel) Assuming that RL is independent of WM, the learned RL value at the end of each block is independent of set size (colors) and capacity (top vs. bottom) but is sensitive to the probability of obtaining 1 point vs. 2 points in correct trials.

Table 1. Experiment 3 Demographics

	HCs	PSZ[a]	*p* Value
n	32	46	
Age, Years, Mean (SD)	37.14 (10.21)	37.81 (8.97)	.76
Education, Years, Mean (SD)			
Participant	15.06 (2.12)	13.27 (2.37)	.001
Maternal	13.75 (2.21)	14.02 (2.93)	.66
Paternal	14.20 (3.68)	13.74 (3.52)	.60
Gender, Male/Female, *n*	21/11	28/18	.58
Race/Ethnicity, *n*			.85
African American	12	18	
White	17	26	
Other	3	2	

HCs, healthy control subjects; PSZ, people with schizophrenia.

[a] Antipsychotic medication regimen (*n*): aripiprazole: 3; clozapine: 20; fluphenazine: 1; haloperidol: 3; lurasidone: 1; olanzapine: 1; quetiapine: 1; risperidone/paliperidone: 6; ziprasidone: 2; multiple antipsychotics: 7; none: 1.

suggest that reward learning deficits in medicated PSZ are likely to result from a failure to represent and compare the expected value of response alternatives. Such representations have been hypothesized to rely on cortical WM function; thus, reductions in WM capacity may drive learning deficits in PSZ (8), with relatively intact learning from striatal RPEs (23). We sought to 1) replicate the observation of selective WM deficits, but not RL deficits, in PSZ during learning (8); 2) show positive evidence during the test phase that RL-dependent learning is unimpaired in PSZ, as predicted from our previous study; and 3) investigate whether interactions of WM and RL are modified in PSZ compared with HCs.

METHODS AND MATERIALS

Experimental Protocol

Experiments 1 and 2 were approved by the Brown University Institutional Review Board and administered to healthy young participants at Brown University. Experiment 3 was approved by the University of Maryland School of Medicine Institutional Review Board and was administered to PSZ and HCs at the Maryland Psychiatric Research Center. Experiment 1 took approximately 1 hour to administer. We conducted experiment 2 in healthy young adults to test whether a shortened version (~ 30 minutes; more appropriate for patient experiments) provided the same power to identify RL and WM effects.

Learning Phase. The experiments used an extension of our Reinforcement Learning/Working Memory task (7). In this protocol (Figure 1A–C), participants used reward feedback to learn which of three actions (key presses with three fingers of the dominant hand) to select in response to different stimuli. There was only one correct action, but the number of points participants could win differed across stimuli; all incorrect actions led to no reward. To manipulate the requirement for capacity-limited and delay-sensitive WM, we varied the set size *ns* (number of image–action associations to learn in a block) across blocks, with new stimuli presented in each new block. Each correct stimulus–action association was assigned a probability *p* of yielding 2 points versus 1 point, and this probability was either high (p = .80), medium (p = .50), or low (p = .20). Stimulus probability assignment was counterbalanced within subjects to ensure equal overall value of different set sizes and motor actions. Depending on the experiment, there were between 10 and 22 blocks of learning for totals of 30 to 75 different stimulus–action associations to be learned.

Test Phase. Following the learning blocks, participants were presented with a surprise test phase. On each test trial, participants were asked to choose which of two images previously encountered in the learning blocks they thought was more rewarding. Participants did not receive feedback during this phase; thus, the ability to select the more rewarding stimulus required having faithfully integrated probabilistic reward magnitude history over learning. Subjects were presented with 156 to 213 pairs during the test phase. Further details of the experiments can be found in the Supplement.

Analysis

Learning Phase. We analyzed the proportion of correct choices as a function of the following variables: set size (number of stimulus images in the block), iteration (how many times the stimulus has been encountered), pcor (number of previous correct choices for the current stimulus), and delay (number of trials since the last correct choice for the current trial's stimulus). See details in the Supplement.

Test Phase. We defined the following characteristics for each image: value (reward history: average of all feedback received for this image), set size, and block (set size and block number of the block in which the stimulus image was encountered). We modeled test performance with a logistic regression with the following key predictors (see Supplement for full details):

ΔQ = value(right) − value(left), assessing value difference effects;

$\Delta ns = ns$(right) − ns(left), assessing whether subjects prefer items that had been encountered in high or low set sizes independent of experienced value, as might be expected if the experience of cognitive effort in high set sizes is aversive;

Mean(ns)*ΔQ, assessing whether value discrimination is stronger or weaker when the items came from relatively high or low set sizes.

RESULTS

Results from the learning phase replicated our previous results, showing that WM and value-based RL both dynamically contribute to learning, even with the presence of probabilistic reward. Indeed, in two separate experiments involving healthy young participants, we observed close-to-optimal learning curves for low set sizes, while performance improved more gradually for higher set sizes even for the equivalent number of iterations per stimulus (Figure 2A). Reaction times decreased with learning and were strongly affected by set size (Figure 2B).

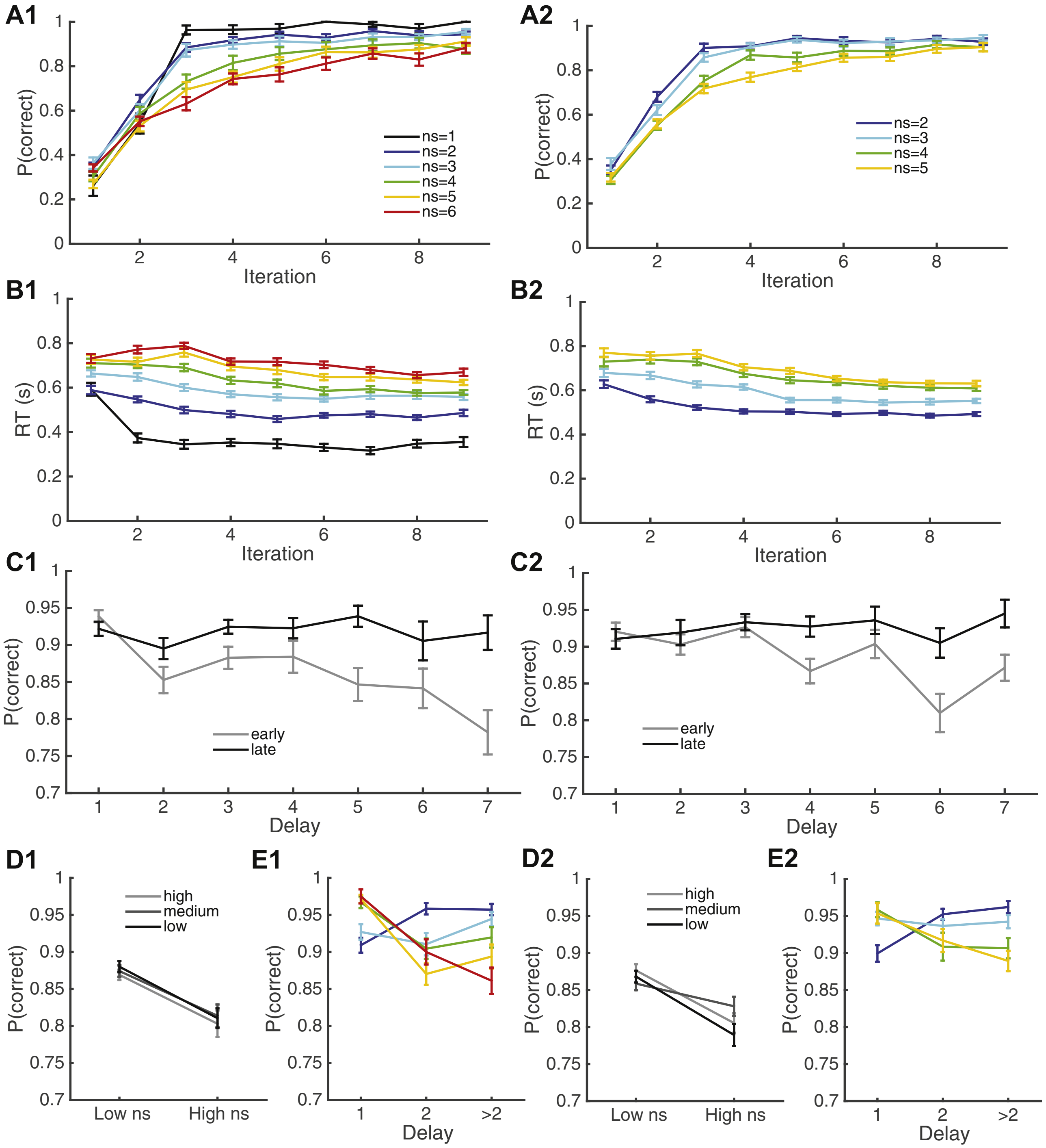

Figure 2. **(A, B)** Learning curves show the proportion of correct trials and mean reaction times (RTs) as a function of the encounter number of each stimulus for different set sizes (*ns*). Left/Right panels show results from experiments 1 and 2, respectively. **(C, E)** Proportion of correct trials as a function of delay (number of trials since correct choice for the current stimulus) for different set sizes or at different learning times (early = up to two prior correct choices; late: final two trials for a given stimulus). **(D)** Performance for stimuli with high, medium, or low probability of reward 2 points vs. 1 point when correct choice is made.

Note that, as elaborated in statistical analyses below, performance decreases in high set sizes were due to a combination of load and the increase in average delay between repeated presentations of the same stimulus (although this delay effect decreased with learning and with lower set sizes, as observed in Figure 2C, E). We found no difference in learning performance for stimuli with a high, medium, or low probability of 2 points versus 1 point (Figure 2D). This can be explained by the fact that reward probability is incidental to the stimulus, but in each case there is always one correct response (see Figure 1 and Methods and Materials).

Experiments 1 and 2

Learning Phase Results. To quantify the effect of RL versus WM, we analyzed learning performance with logistic regression on trial-by-trial data, allowing us to parse out effects of delay from those of set size. In a first analysis, including only set size, number of previous correct choices, and delay as predictors, in both experiments we found strong effects of all three factors: worse performance with higher set size (experiment 1: $t_{27} = -5.3$, $p < 10^{-4}$; experiment 2: $t_{50} = -2.8$, $p = .007$), worse performance with higher delay (experiment 1: $t_{27} = -2.8$, $p = .009$; experiment 2: $t_{50} = -2.9$, $p = .005$), and better performance with increasing previous correct choices (experiment 1: $t_{27} = 15.9$, $p < 10^{-4}$; experiment 2: $t_{50} = 7.5$, $p < 10^{-4}$). Follow-up analysis with interaction terms replicated previously published results (Figure 3) showing that delay effects were stronger in higher set sizes and decreased with iterations (both $ps < 10^{-4}$, $t > 7.5$ for experiment 1; $ts = -3.1$ and 2.3 and $ps = .002$ and .02, respectively, for experiment 2), with the interaction between set size and iterations not reaching significance (experiment 1: $p = .10$, $t = 1.7$; experiment 2: $p = .13$, $t = 1.6$).

Taken together, these results confirm that both WM and RL contributed to learning in this task and hint at a possible shift from capacity-limited, but fast, WM to incremental RL after increasing exposure, with a weakening of the effects of delay and set size with iterations. The slightly weaker effects observed in experiment 2 might be due to a smaller spread of set sizes (2–5 instead of 1–6) and about half the number of trials, weakening the inference of the logistic regression (Figure 3). However, because the effects were very comparable across the two experiments, we next report test phase results pooled across both (but see figures for results within each experiment).

Test Phase Results. We first confirmed that participants had indeed encoded the reward values; in a logistic regression analysis, participants were significantly more likely to select the higher value image (Figure 4, left; $t_{66} = 3.0$, $p = .003$), showing sensitivity to the value difference between two images.

We next asked whether sensitivity to value difference depended on whether the stimuli had been learned in high or low set size blocks. Surprisingly, we found that value discrimination was enhanced when the items were learned in high set size blocks rather than low set size blocks ($t_{66} = 2.3$, $p = .03$). In particular, when we analyzed choice within trials where both images came from a high set size block ($ns > 4$) and compared choice on trials where both images came from a low set size block ($ns < 4$), we found that participants were sensitive to value differences in both subsets (Figure 4, right; both $ts > 3.3$, $ps \leq .001$), but significantly more so in high set sizes ($t = 2.7$, $p = .008$). This result indicates that the value learning process is different when WM is differently engaged, hinting at a potential interaction between the WM and RL systems (see below).

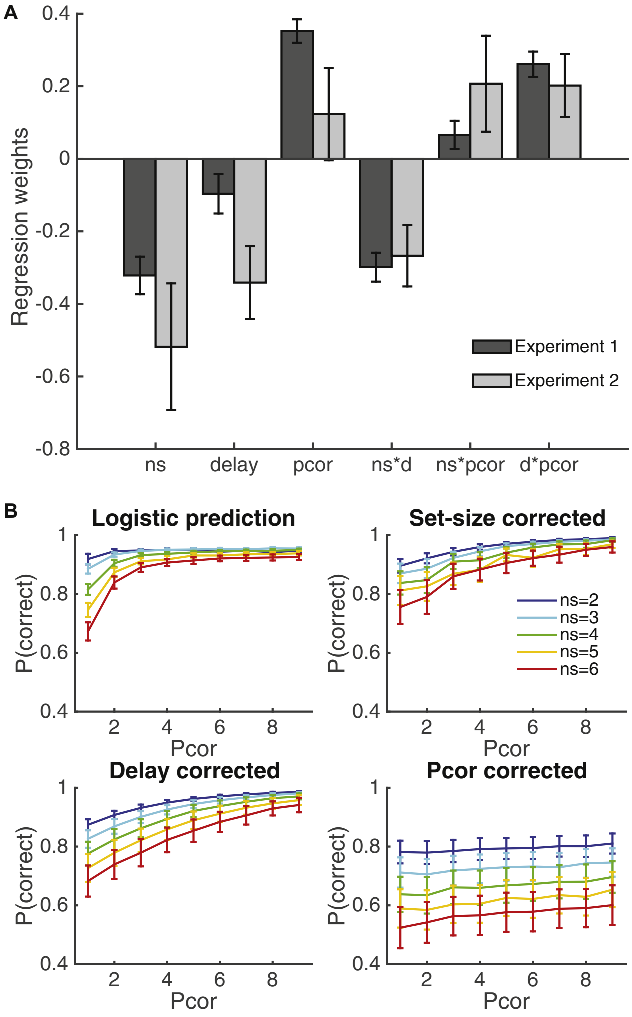

Figure 3. Learning phase. **(A)** Results from the logistic show consistent effects of set size, delay, number of previous correct choices (Pcor), and interactions (except set size with pcor). Error bars indicate standard error of the mean. **(B)** Experiment 1 logistic regression predictions (top left panel) show set size and pcor effect within trials with at least one previous correct choice for the current stimulus. Logistic predictions when correcting for set size or delay still show a remaining effect of set size, indicating that both factors play an important role in explaining the slower learning in higher set sizes.

Finally, participants were significantly more likely to select an item from a low set size block than from a high set size block (Figure 4, left; $t_{66} = -4.4$, $p < 10^{-4}$). This result is consistent with other studies indicating that cognitive effort associated with WM demand or response conflict confers a cost (24–26) that translates into reduced effective value learning (27).

Experiment 3

Learning Phase Results. We next used this task to investigate learning impairments in medicated PSZ. PSZ had fewer correct responses than HCs ($t_{77} = 2.7$, $p = .007$; Cohen's

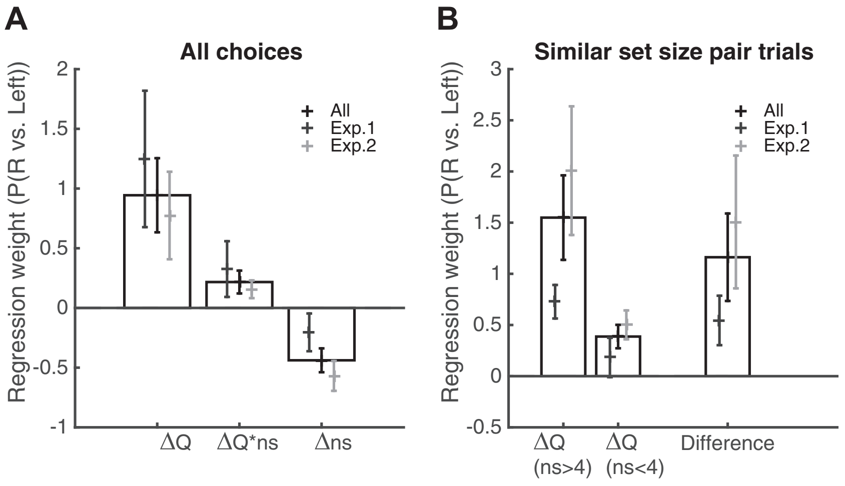

Figure 4. Test phase results. **(A)** We analyzed choice of the right (R) vs. left image in the test phase as a function of the value difference ΔQ = value(right) − value(left), the set size difference Δns, and $\Delta Q{*}ns$, the interaction of the mean set size of the two images with the value difference, as well as other regressors of noninterest. We found a significant effect of all three factors across both experiments (Exp). **(B)** The effect of value difference is significantly stronger in high set sizes than in low set sizes, indicating that reinforcement learning was more efficient under high load, thereby highlighting an interaction of reinforcement learning with working memory.

d = 0.63), and this was true in all set sizes 3 to 5 ($ts > 2.2$, $ps < .03$; Figure 5, left), with only marginal deficits in set size 2 (t_{77} = 1.4, $p > .10$). Based on our previous report, we hypothesized that PSZ would show reduced WM capacity for guiding learning (8) and, hence, would show greater differences in performance between sequentially adjacent set sizes once they were above capacity. Figure 5 (top right) compares performances for sequentially adjacent set sizes. We observed that performance in HCs was not significantly different between set sizes 2 and 3 (t_{31} = 1.3, $p > .10$), whereas there was a strong decrease in PSZ performance between these sets (t_{46} = 5.7, $p < 10^{-4}$); the difference between the two groups was significant (t_{77} = 2.6, p = .01). HCs' performance instead decreased between set sizes 3 and 4 (t_{31} = 4.0, p = .0003), supporting the interpretation that they had larger capacity [between 3 and 4 as reported in earlier studies (7,8)]. There was no other difference in the change in performance with set size between the two groups. Thus, the main finding is that HCs treat set sizes 2 and 3 as equivalent and suffer further decrements in performance with each additional increase in load, whereas PSZ already suffer from a difference in load between 2 and 3.

The logistic regression analysis (Figure 5, bottom right) confirmed previous observations (including those in experiments 1 and 2) that probability of correct choices decreased with set size and delay and increased with the number of

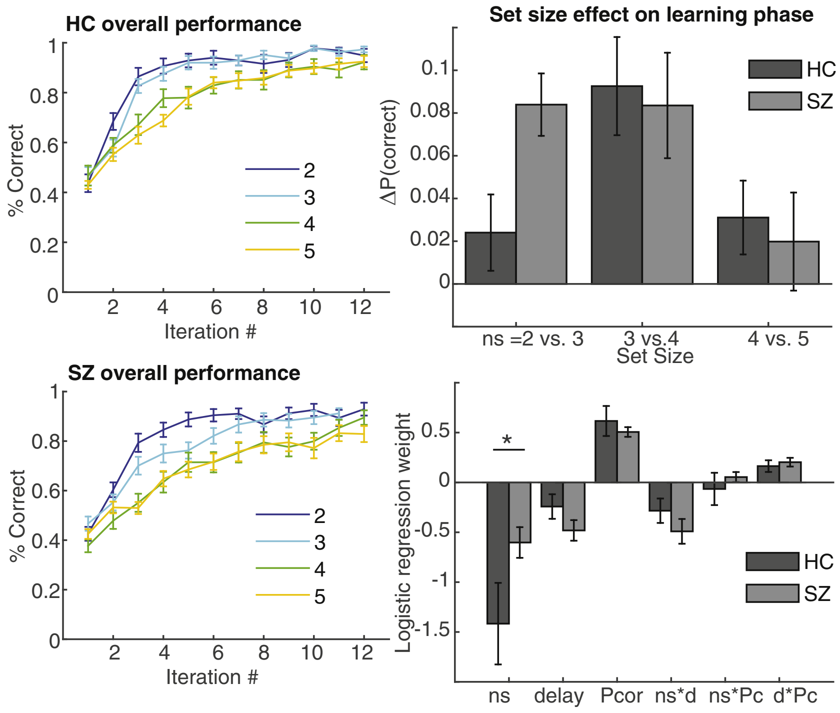

Figure 5. Schizophrenia learning phase results replicated our previous finding that working memory contributes to learning impairment. (Left panel) Learning curves (see Figure 2) show slower learning for people with schizophrenia (SZ) than for healthy control subjects (HC). (Top right panel) Change in performance from set size 2 to 3 is significantly higher in SZ than in HC. The HC pattern matches a capacity 3 model simulation (Figure 1E), while the SZ pattern matches a mixture of capacity 2 and capacity 3 model simulation. (Bottom right panel) Logistic regression analysis shows a difference in the set size effect only between groups, implicating the working memory mechanism.

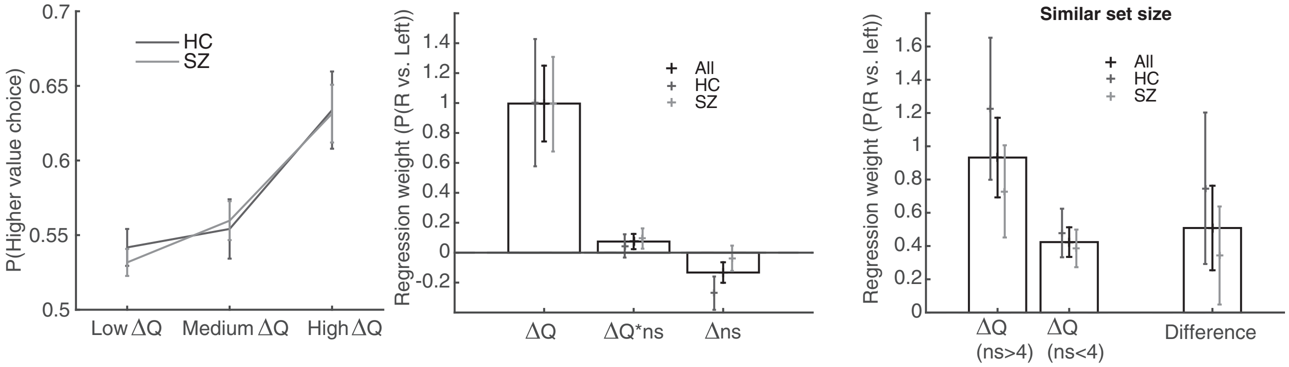

Figure 6. Schizophrenia test phase supported our prediction that reinforcement learning–dependent value learning is unimpaired in people with schizophrenia (SZ). (Left panel) Proportion of higher value choices increases with the value difference between the two items in a trial (grouped in tertiles based on absolute value difference); however, there was no difference between SZ and healthy control subjects (HC). (Middle panel) Logistic regression analysis of the test phase supported our finding that SZ and HC were equally sensitive to value difference. We found an effort effect in HCs but not in SZs. (Right panel) Both groups were more sensitive to value difference in high set sizes (*ns*) than in low set sizes, supporting our previous result. R, right.

previous correct choices (*t*s$_{77}$ $>$ 4.7, *p*s $<$ 10^{-4}). There were also significant interactions of delay with set size and number of previous correct choices (*t*s$_{77}$ $>$ 4.5, *p*s $<$ 10^{-4}), but there was not a significant interaction between set size and number of previous correct choices (t_{77} = 0.05, nonsignificant). The only significant difference between groups was observed for the set size effect (t_{76} = 2.1, p = .04; all other *t*s $<$ 1.45), indicating a weaker effect of set size in PSZ than in HCs. This result is consistent with the notion that PSZ performance was less reliant on WM for guiding learning, indicating that PSZ exhibit reduced effects of manipulations that load on WM. The result further supports the previously published finding that PSZ exhibited a deficit in performance already at *ns* = 3 and therefore show less influence of further increases in load. Moreover, as reported earlier, incremental RL processes appeared to be intact, as suggested by the failure to find a significant difference between PSZ and HCs in the effect of number of previous correct iterations.

Test Phase Results. In contrast to the robust learning phase deficits, PSZ exhibited an identical ability to select stimuli having larger probabilistic reward values (Figure 6, left). Specifically, for each group, performance for each tertile of value difference (low, medium, or high) was significantly better than chance (*t*s $>$ 2.7, *p*s $<$ .01), and performance increased with value difference (high vs. medium or low, all *t*s $>$ 3.5, p $<$ .001). Furthermore, in each tertile, performance was indistinguishable between the two groups (all *t*s $<$ 0.54).

Test phase logistic regression analysis confirmed our previous observation; across the whole group, the effect of value difference on choice was significant (Figure 6, middle; ΔQ t_{70} = 3.9, p = .0002), and there was no difference between the two groups (t = 0.25, nonsignificant). Next, we investigated whether value learning changed with set size, as found in the previous two experiments. Although the effect did not reach significance across the whole group in the analysis including all trials (t = 1.46, p = .15), it was significant in the more targeted analysis comparing sensitivity to value difference within high set size pairs compared with low set size pairs (Figure 6, right; t_{70} = 2.0, p $<$ .05), supporting our previous observation. There was no difference between PSZ and HCs (all *t*s $<$ 1.5, *p*s $>$ .10). Finally, we investigated the previously found effort effect, whereby young healthy participants were more likely to select an image from a low set size block than from a high set size block. We replicated this effect in HCs (t_{28} = 2.4, p = .02), but interestingly, we did not observe a similar effect in PSZ (t_{41} = 0.44). However, the difference between groups was not significant (t = 1.5, p = .14). We did not find any relation between either test or learning phase performance and symptom ratings, neuropsychological performance, or antipsychotic dose.

DISCUSSION

Findings from our new protocol extend those from our previously developed learning task, enabling us to identify separable contributions of WM and RL to learning and highlighting the role of WM in apparently model-free learning behavior (see Supplement). Indeed, in all three current experiments, learning performance was sensitive to load and delay, hallmarks of WM use, as well as to reward history, a hallmark of RL. Moreover, WM effects decreased as learning progressed, supporting prior computational modeling results suggesting a transition from WM to RL (7). Our new protocol also provides additional sensitivity to probabilistic value learning within the RL system and, more explicitly, reveals interactions between WM and RL. The enhanced design was able to replicate our previous finding that impaired WM in PSZ substantially contributes to PSZ poor instrumental learning performance. Strikingly, despite marked learning impairments driven by putative WM processes, the test phase results more definitively show that PSZ successfully integrated reward values—under the purview of RL. Overall, the consistent results reported across the three experiments presented here highlight a significant benefit of designing and analyzing experiments within a computational framework that disentangles contributions to learning.

In addition to including probabilistic rewards, one of the advancements of this task was to include a surprise test phase in which participants were able to reliably select the images that had been most rewarded. We argue that this ability reflects the expression of RL processes. Indeed, participants were exposed to a large number of images during the learning phase (78 and 39 in experiments 1 and 2, respectively), far exceeding the capacity of WM (7,8,28). Furthermore, participants did not need to explicitly integrate the value of each image during learning; indeed, the number of points they received per stimulus was not controllable, and participants were unaware of the upcoming test phase, so any value learning occurred incidentally. Finally, the type of choices assessed during the test phase is similar to that in older tasks showing sensitivity to probabilistic value integration (19,21,29–31). Indeed, a recent study assessing model-based and model-free RL revealed dissociable prefrontal cortex and striatal genotypes that relate to model-based function during learning and probabilistic integration of value learning assessed during test, respectively (3). While these prior tasks demonstrated effects of striatal dopamine and individual differences thereof on sensitivity to learning from positive versus negative outcomes, future work will need to assess whether similar biases are induced by manipulations in our analogous measure of biased learning during the test phase.

In addition to improving sensitivity to RL while retaining sensitivity to WM, our new protocol allows us to investigate their interaction. We observed two interesting interactions between the two systems. First, we observed a cognitive effort effect on RL; during the test phase, participants were more likely to select an image that had been encountered in a low set size block than in a high set size bock independent of the difference in value between the two images. Cognitive control is effortful and may be aversive (24–26), and conflict, which requires cognitive control to resolve, is aversive and leads to reductions in learned value (27,32,33). This notion is consistent with our observation here that effective values are reduced for items that had been encountered under high WM load.

Second, we also observed a more counterintuitive interaction, whereby participants exhibited enhanced ability to discriminate objective differences in value when the two items had been learned in high set sizes (i.e., when learning was more difficult) than when they had been learned in low set sizes. This result highlights an interference of WM computations into RL computations. We propose that this interaction can be accounted for by a competitive or cooperative computational mechanism linking WM with RL. According to the competitive account, successful engagement of WM in low set sizes inhibits the RL system from accumulating values and, hence, hindering subsequent value discrimination. Alternatively, a cooperative account assumes that RL operates regardless of load but that expectations in WM provide input to the RL system so that prediction errors are reduced when WM is successful (i.e., in low set sizes). As such, positive RPEs would be blunted with a working WM–RL interaction, leading to reduced integration of value in the RL system. Future work may be able to disentangle these competing explanations with imaging. In either case, our protocol allowed us to show that RL and WM do not operate separately but that WM interferes with RL computations.

Disentangling the role of multiple systems in learning is crucial to link individual differences in behavior to the neural mechanisms supporting them. This is particularly true in psychiatric research; many psychiatric diseases include learning impairments, and knowing whether such impairments are more likely related to the striatal-dopaminergic integration of reward and punishment over time, or to WM use, would be an important step toward a better understanding of the neural systems implicated in the disease. Here, we exemplify this with the case of schizophrenia. Learning impairments have been broadly observed in PSZ, but the nature of these impairments remains unclear (22), with conflicting findings across studies at the behavioral level [impairments in some learning situations but not in others (34,35)] and at the neural level [identifying different striatal signals vs. controls (36–38)]. In a previously published study (8), we found that overall learning impairments in PSZ were entirely explained by WM contributions to learning, with no difference in the RL contributions between PSZ and controls. However, our initial study was less sensitive to RL than to WM because of the use of fully deterministic stimulus–action–outcome contingencies. Here, we provide a complete conceptual replication of our previous finding of WM impairments explaining poorer learning during the initial learning phase. This is particularly noteworthy in that we used probabilistic, as opposed to deterministic, feedback and examined different set size ranges across experiments, suggesting that this finding is likely quite robust and reliable. With the addition of the test phase, we more explicitly showed that PSZ possess fully intact ability to accumulate statistics of probabilistic values because their ability to discriminate items based on these learned values was indistinguishable from HCs. Given that PSZ typically demonstrate impairments relative to HCs in effortful cognitive tasks, the fact that we have now seen fully normal performance levels in striatal RL across two independent experiments is a noteworthy example of the value of computational approaches. Our results were not linked to medication dosage and did not provide insight as to whether specific symptoms (beyond cognitive symptoms), in particular negative symptoms, were linked to distinct contributions to learning (see Supplement for additional results and discussion).

Conclusions

We introduced a protocol designed to disentangle the role of RL and WM in instrumental performance and showed that this protocol is sensitive to individual differences in both processes and allows us to investigate their interaction. Behavioral results showed that the two processes compete for choice during learning, and at a deeper level, as they perform their computations. Specifically, we hypothesized that WM contributes expectations to the computation of RL RPEs, thereby ironically weakening learning in the RL substrate. We demonstrated the usefulness of our protocol in an experiment comparing learning in HCs and PSZ, confirming that learning impairments in PSZ are due to WM while RL is fully spared. More generally, we hope that this protocol can get us closer to underlying neural mechanisms supporting human learning and, thus, further our understanding of healthy learning as well as learning impairments in different clinical populations.

ACKNOWLEDGMENTS AND DISCLOSURES

This work supported by the National Institute of Mental Health (Grant No. RO1 MH080066 to JMG and MJF) and Grant No. NSF1460604 (to MJF and AGEC). MAA is supported by the National Health and Medical Research Council (Grant No. APP1090716).

We thank Julie Helmers for her help in data collection for experiments 1 and 2 and thank Sharon August and Leeka Hubzin for their work on the experiment conducted at the Maryland Psychiatric Research Center.

AGEC, MJF, JAW, and JMG report that they perform consulting for Hoffman La Roche. JMG has also consulted for Takeda and Lundback and receives royalty payments from the Brief Assessment of Cognition in Schizophrenia. JAW also consults for NCT Holdings. The current experiments were not related to any consulting activity. MAA reports no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Helen Wills Neuroscience Institute (AGEC), Department of Psychology, University of California, Berkeley, Berkeley, California; Maryland Psychiatric Research Center (MAA, JAW, JMG), Department of Psychiatry, University of Maryland School of Medicine, Baltimore, Maryland; Curtin Health Innovation Research Institute (MAA), School of Public Health, Curtin University, Perth, Western Australia, Australia; and Brown Institute for Brain Sciences (MJF), Department of Cognitive Linguistic and Psychological Science, Brown University, Providence, Rhode Island.

Address correspondence to Anne G.E. Collins, Ph.D., Department of Psychology, University of California, Berkeley, 3210 Tolman Hall, Berkeley, CA 94720; E-mail: annecollins@berkeley.edu.

Received Dec 13, 2016; revised May 9, 2017; accepted May 10, 2017.

Supplementary material cited in this article is available online at http://dx.doi.org/10.1016/j.biopsych.2017.05.017.

REFERENCES

1. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011): Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204–1215.
2. Doll BB, Duncan KD, Simon DA, Shohamy D, Daw ND (2015): Model-based choices involve prospective neural activity. Nat Neurosci 18:767–772.
3. Doll BB, Bath KG, Daw ND, Frank MJ (2016): Variability in dopamine genes dissociates model-based and model-free reinforcement learning. J Neurosci 36:1211–1222.
4. Wunderlich K, Smittenaar P, Dolan RJ (2012): Dopamine enhances model-based over model-free choice behavior. Neuron 75:418–424.
5. Gläscher J, Daw N, Dayan P, O'Doherty JP (2010): States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66:585–595.
6. Cushman F, Morris A (2015): Habitual control of goal selection in humans. Proc Natl Acad Sci U S A 112:13817–13822.
7. Collins AGE, Frank MJ (2012): How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. Eur J Neurosci 35:1024–1035.
8. Collins AGE, Brown JK, Gold JM, Waltz JA, Frank MJ (2014): Working memory contributions to reinforcement learning impairments in schizophrenia. J Neurosci 34:13747–13756.
9. Klingberg T (2010): Training and plasticity of working memory. Trends Cogn Sci 14:317–324.
10. Baddeley A (2003): Working memory: Looking back and looking forward. Nat Rev Neurosci 4:829–839.
11. Baddeley A (2012): Working memory: Theories, models, and controversies. Annu Rev Psychol 63:1–29.
12. D'Esposito M, Postle BR (2015): The cognitive neuroscience of working memory. Annu Rev Psychol 66:115–142.
13. Sutton RS, Barto AG (1998): Reinforcement learning: An introduction. IEEE Trans Neural Netw 9:1054.
14. Tan H-Y, Chen Q, Goldberg TE, Mattay VS, Meyer-Lindenberg A, Weinberger DR, *et al.* (2007): Catechol-*O*-methyltransferase Val158Met modulation of prefrontal-parietal-striatal brain systems during arithmetic and temporal transformations in working memory. J Neurosci 27:13393–13401.
15. Cohen JR, Gallen CL, Jacobs EG, Lee TG, D'Esposito M (2014): Quantifying the reconfiguration of intrinsic networks during working memory. PLoS One 9:e106636.
16. McNab F, Klingberg T (2008): Prefrontal cortex and basal ganglia control access to working memory. Nat Neurosci 11:103–107.
17. Montague PR, Dayan P, Sejnowski TJ (1996): A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16:1936–1947.
18. Collins AGE, Frank MJ (2014): Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychol Rev 121:337–366.
19. Frank MJ, Seeberger LC, O'Reilly RC (2004): By carrot or by stick: Cognitive reinforcement learning in parkinsonism. Science 306:1940–1943.
20. Pessiglione M, Petrovic P, Daunizeau J, Palminteri S, Dolan RJ, Frith CD (2008): Subliminal instrumental conditioning demonstrated in the human brain. Neuron 59:561–567.
21. Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007): Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl Acad Sci U S A 104:16311–16316.
22. Deserno L, Boehme R, Heinz A, Schlagenhauf F (2013): Reinforcement learning and dopamine in schizophrenia: Dimensions of symptoms or specific features of a disease group? Front Psychiatry 4:172.
23. Gold JM, Strauss GP, Waltz JA, Robinson BM, Brown JK, Frank MJ (2013): Negative symptoms of schizophrenia are associated with abnormal effort–cost computations. Biol Psychiatry 74:130–136.
24. Kool W, Botvinick M (2014): A labor/leisure tradeoff in cognitive control. J Exp Psychol Gen 143:131–141.
25. Westbrook A, Braver TS (2015): Cognitive effort: A neuroeconomic approach. Cogn Affect Behav Neurosci 15:395–415.
26. Shenhav A, Botvinick MM, Cohen JD (2013): The expected value of control: An integrative theory of anterior cingulate cortex function. Neuron 79:217–240.
27. Cavanagh JF, Masters SE, Bath K, Frank MJ (2014): Conflict acts as an implicit cost in reinforcement learning. Nat Commun 5:5394.
28. Cowan N (2010): The magical mystery four: How is working memory capacity limited, and why? Curr Dir Psychol Sci 19:51–57.
29. Doll BB, Hutchison KE, Frank MJ (2011): Dopaminergic genes predict individual differences in susceptibility to confirmation bias. J Neurosci 31:6188–6198.
30. Cockburn J, Collins AGE, Frank MJ (2014): A reinforcement learning mechanism responsible for the valuation of free choice. Neuron 83:551–557.
31. Cox SML, Frank MJ, Larcher K, Fellows LK, Clark CA, Leyton M, *et al.* (2015): Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes. NeuroImage 109:95–101.
32. Dreisbach G, Fischer R (2012): Conflicts as aversive signals. Brain Cogn 78:94–98.
33. Fritz J, Dreisbach G (2013): Conflicts as aversive signals: Conflict priming increases negative judgments for neutral stimuli. Cogn Affect Behav Neurosci 13:311–317.
34. Gold JM, Hahn B, Strauss GP, Waltz JA (2009): Turning it upside down: Areas of preserved cognitive function in schizophrenia. Neuropsychol Rev 19:294–311.
35. Heerey EA, Bell-Warren KR, Gold JM (2008): Decision-making impairments in the context of intact reward sensitivity in schizophrenia. Biol Psychiatry 64:62–69.
36. Waltz JA, Kasanova Z, Ross TJ, Salmeron BJ, McMahon RP, Gold JM, *et al.* (2013): The roles of reward, default, and executive control networks in set-shifting impairments in schizophrenia. PLoS One 8:e57257.
37. Schlagenhauf F, Huys QJM, Deserno L, Rapp M a, Beck A, Heinze H-J, *et al.* (2014): Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. NeuroImage 89:171–180.
38. Dowd EC, Frank MJ, Collins A, Gold JM, Barch DM (2016): Probabilistic reinforcement learning in patients with schizophrenia: Relationships to anhedonia and avolition. Biol Psychiatry Cogn Neurosci Neuroimaging 1:460–473.