Published in final edited form as:
Biol Psychiatry. 2017 September 15; 82(6): 431–439. doi:10.1016/j.biopsych.2017.05.017.

Interactions between working memory, reinforcement learning and effort in value-based choice: a new paradigm and selective deficits in schizophrenia

Anne GE Collins[*],
UC Berkeley, Dept of Psychology, HWNI

Matthew A Albrecht,
Maryland Psychiatric Research Center, Department of Psychiatry, School of Medicine, University of Maryland, Baltimore, Maryland, USA; Curtin University, School of Public Health, Curtin Health Innovation Research Institute, Perth, Western Australia, Australia

James A. Waltz,
Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine. Baltimore, Maryland, USA

James M. Gold, and
Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine. Baltimore, Maryland, USA

Michael J Frank
Brown University, Cognitive Linguistic and Psychological Science department, Brown Institute for Brain Sciences

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 design tasks that allow such a separable identification of processes, and infer their contributions in individuals.

[*]corresponding author: annecollins@berkeley.edu; (510) 664-7146; UC Berkeley Dept of Psychology, 3210 Tolman Hall, Berkeley CA 9420.

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

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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 1-2, we test this protocol with healthy young adults (n=29 and n=52). In experiment 3, we use it to investigate learning deficits in medicated individuals with schizophrenia (n=49 patients, n=32 controls).

Results—Experiments 1-2 established WM and RL contributions to learning, evidenced by parametric modulations of choice by load and delay, and reward history, respectively. It also showed interactions between WM and RL, where RL was enhanced under high WM load. Moreover, we observed a cost of mental effort, 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 to controls.

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

Keywords

Working memory; Reinforcement learning; effort; schizophrenia; computational modeling; decision-making

Intro

Multiple neurocognitive systems interact to support various forms of learning, each with their 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 vs. stimulus-response habit formation during sequential multi-stage reinforcement learning (1–6). However, these processes can interact, and moreover, they can themselves be subdivided into multiple systems (e.g., planning involves working memory, accurate representation of environmental contingencies, guided strategic search through such contingencies to determine a desired course of action, etc.).

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 working memory and reinforcement learning (7,8). We refer to working memory (WM) 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). Working memory 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 (RL) 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/PFC function (14–16), and RL relying on phasic dopaminergic signals conveying reward prediction errors that modulate 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 (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 (Fig. 1A-C) was motivated by prior modeling of WM and RL contributions to learning (Fig. 1D-E), and extended our previous design with two new features. First, we added probabilistic variation in reward magnitudes (1 vs. 2 points) across stimuli (Fig. 1 A-B), and second, we added a subsequent surprise test phase (Fig. 1 C), 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 (Fig. 1D). Furthermore, this improved task allows us to investigate whether WM demands during learning also influence the degree of value learning (Fig. 1E). Such interactions would motivate refinement of existing computational models which assume that RL and WM processes proceed independently and only compete 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 controls (HC) matched on important demographic variables. The literature remains unclear as to the specific nature of learning impairments in PSZ (22). Indeed, recent studies 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 of learning deficits in PSZ (8), with relatively intact learning from striatal reward prediction errors (23). We sought to 1) replicate the observation of selective WM but not RL deficits in PSZ during learning (8), 2) show positive evidence in 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 to controls.

Methods

Experimental protocol

Experiments 1 and 2 were approved by the Brown University institutional review board (IRB), and administered to healthy young participants at Brown University. Experiment 3 was approved by the University of Maryland School of Medicine IRB, and administered to PSZ and HC at the Maryland Psychiatric Research Center (MPRC). Experiment 1 took approximately 1 hour to administer. We conducted Experiment 2 in healthy young adults to test whether a shortened version (ca. 30 min; more appropriate for patient experiments) still provided the same power to identify RL and WM effects.

Learning phase—The experiments used an extension of our Reinforcement Learning/ Working Memory (RLWM) task (7). In this protocol (Fig. 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 lead 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 vs. 1 point, and this probability was either high ($p=0.8$), medium ($p=0.5$), or low ($p=0.2$). 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 10-22 blocks of learning, for totals of 30-75 different stimulus-action associations to be learned.

Testing 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 156-213 pairs in the test phase. Further details of the experiments can be found in the Supplementary Information (SI).

Analysis

Learning phase—We analyzed the proportion of correct choices as a function of the 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 SI.

Test phase—We defined for each image the following characteristics: *value* (reward history: average of all feedback received for this image), *set size*, and *block* (the 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.

Δns $= ns(right)-ns(left)$, assessing whether subjects prefer items that had been encountered in high or low set sizes independently 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 working memory 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 (Fig. 2A). Reaction times decreased with learning and were strongly affected by set size (Fig. 2B). 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 (though this delay effect decreased with learning and with lower set sizes, as observed in Fig. 2C and 2E). We found no difference in learning performance for stimuli with a high, medium, or low probability of 2 points vs. 1 point (Fig. 2 D). 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 Fig. 1 and Methods).

Learning phase results—To quantify the effect of reinforcement learning vs. working memory, 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, we found in both experiments strong effects of all three factors: worse performance with higher set size (Exp. 1: t(27)=-5.3, p<10-4; Exp. 2: t(50)=-2.8, p=0.007), worse performance with higher delay (Exp. 1: t(27)=-2.8, p=0.009; Exp. 2: t(50)=-2.9, p=0.005), and better performance with increasing previous correct choices (Exp. 1: t(27)=15.9, p<10-4; Exp. 2 t(50)=7.5, p<10-4). Follow-up analysis with interaction terms replicated previous published results (Fig. 2F) showing that delay effects were stronger in higher set sizes and decreased with iterations (both p's<10^{-4}, t>7.5 for Exp1; t= -3.1 and 2.3, p=.002, .02 respectively for Exp. 2), and the interaction between set size and iterations not reaching significance (Exp.1 p=0.1, t=1.7; Exp2 p=0.13, 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 Exp. 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. 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 (Fig. 3 left; t(66)=3.0, p=0.003), 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 rather than low set size blocks (t(66)=2.3, p=0.03). 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 (Fig. 3 right both t>3.3, p≤0.001), but significantly more so in high set sizes (t=2.7, p=0.008). This result indicates that the value learning process is different when working memory is differently engaged, hinting at a potential interaction between the working memory and reinforcement learning systems (see below).

Finally, participants were significantly more likely to select an item from a low set size block than a high set size block (Fig. 3, left; t(66)=-4.4, p<10-4). This result is consistent with other studies indicating that cognitive effort associated with working memory 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 people with schizophrenia. PSZ had fewer correct responses than healthy controls (t(77)=2.7, p=0.007; Cohen's d = 0.63), and this was true in all set sizes 3-5 (ts>2.2, ps<0.03; Fig. 5, left), with only marginal deficits in set size 2 (t(77)=1.4, p>.1). 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. Fig. 5 (top right) compares performances for sequentially-adjacent set sizes. We observed that performance in healthy controls was not significantly different between set sizes 2 and 3 (t(31)=1.3, p>.1), 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=0.01). Controls' performance instead decreased between set sizes 3 and 4 (t(31)=4.0, p=0.0003), 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 HC 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 (Fig. 5, bottom right) confirmed previous observations (including those in Experiments 1-2) that probability of correct choices decreased with set size and delay, and increased with the number of previous correct choices (ts(77)>4.7, ps<10^{-4}). There were also significant interactions of delay with set size and number of previous correct choices (ts(77)>4.5, ps<10^{-4}), but there was not a significant interaction between set size and number of previous correct choices (t(77)=.05,ns.). The only significant difference between groups was observed for the set size effect (t(76)=2.1, p=0.04; all other ts<1.45), indicating a weaker effect of set size in PSZ than controls. 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 controls in the effect of number of previous correct iterations (ns).

Test phase results—In contrast to the robust learning phase deficits, PSZ exhibited an identical ability to select stimuli having larger probabilistic reward values (Fig. 6, left). Specifically, for each group, performance for each tertile of value difference (low, medium and high) was significantly better than chance (ts>2.7, ps<0.01), and performance increased with value difference (high vs. medium or low, all ts>3.5, p<0.001). 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 (Fig. 6 middle, ΔQ t(70)=3.9, p=0.0002), and there was no difference between the two groups (t=0.25, ns). 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=0.15), it was significant in the more targeted analysis comparing sensitivity to value difference within high set size pairs compared to low set size pairs (Fig. 6, right; t(70)=2, p<0.05), supporting our previous observation. There was no difference between PSZ and controls (all t's<1.5, p's>.1). Finally, we investigated the previously found “effort” effect, whereby young healthy participants were more likely to select an image from a low set size than high set size block. We replicated this effect in healthy controls (t(28)=2.4, p=0.02), 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=0.14). We did not find any relation between either test or learning phase performance with 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 SI). 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 people with schizophrenia (PSZ) substantially contributes to their 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 in 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, thus any value learning occurred incidentally. Finally, the type of choices assessed in the test phase are 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 PFC 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 vs. negative outcomes; future work will need to assess whether similar biases are induced by manipulations in our analogous measure of biased learning in 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: in the test phase, participants were more likely to select an image which had been encountered in a low set size block than in a high set size bock, independently from 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 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 working memory 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 (with impairments in some learning situations but not others (34,35), and at the neural level (identifying different striatal signals compared to 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 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 in 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, as their ability to discriminate items based on these learned values was indistinguishable from controls. Given that PSZ typically demonstrate impairments relative to controls 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 reinforcement learning and working memory on instrumental performance, and showed that this protocol is sensitive to individual differences in both processes, and lets us 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 reward prediction errors, thus ironically weakening learning in the RL substrate. We demonstrated the usefulness of our protocol in an experiment comparing learning in healthy controls and people with schizophrenia, 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 the underlying neural mechanisms supporting human learning, and thus further our understanding of healthy learning as well as learning impairments in different clinical populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work supported by NIMH RO1 MH080066. MA is supported by NHMRC APP1090716.

The authors thank Julie Helmers for her help in data collection for experiments 1 and 2, and Sharon August and Leeka Hubzin for their work on the experiment conducted at MPRC.

References

1. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. Neuron Elsevier Inc. 2011 Mar 24; 69(6):1204–15.
2. Doll, BB., Duncan, KD., Simon, Da, Shohamy, D., Daw, ND. Nat Neurosci. Nature Publishing Group; 2015 Mar 23. Model-based choices involve prospective neural activity; p. 1-9.
3. Doll BB, Bath KG, Daw ND, Frank MJ. Variability in Dopamine Genes Dissociates Model-Based and Model-Free Reinforcement Learning. J Neurosci. 2016; 36(4):1211–22. [PubMed: 26818509]
4. Wunderlich K, Smittenaar P, Dolan RJ. Dopamine enhances model-based over model-free choice behavior. Neuron. 2012 Aug 9; 75(3):418–24. [PubMed: 22884326]
5. Gläscher, J., Daw, N., Dayan, P., O'Doherty, JP. Neuron. Vol. 66. Elsevier Ltd; 2010 May 27. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning; p. 585-95.
6. Cushman F, Morris A. Habitual control of goal selection in humans. Proc Natl Acad Sci. 2015; 112(45):201506367.
7. Collins AGE, Frank MJ. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. Eur J Neurosci. 2012 Apr; 35(7):1024–35. [PubMed: 22487033]
8. Collins, aGE., Brown, JK., Gold, JM., Waltz, Ja, Frank, MJ. Working Memory Contributions to Reinforcement Learning Impairments in Schizophrenia. J Neurosci. 2014 Oct 8; 34(41):13747–56. [PubMed: 25297101]
9. Klingberg, T. Trends Cogn Sci. Vol. 14. Elsevier Ltd; 2010 Jul. Training and plasticity of working memory; p. 317-24.
10. Baddeley A. Working memory: looking back and looking forward. Nat Rev Neurosci. 2003 Oct; 4(10):829–39. [PubMed: 14523382]
11. Baddeley A. Working memory: theories, models, and controversies. Annu Rev Psychol. 2012 Jan. 63:1–29. [PubMed: 21961947]
12. D'Esposito M, Postle BR. The Cognitive Neuroscience of Working Memory. Annu Rev Psychol. 2015; 66(1):115–42. [PubMed: 25251486]
13. Sutton RS, Barto aG. Reinforcement Learning: An Introduction. IEEE Trans Neural Networks. 1998 Sep; 9(5):1054–1054.
14. Tan HY, Chen Q, Goldberg TE, Mattay VS, Meyer-Lindenberg A, Weinberger DR, et al. Catechol-O-methyltransferase Val158Met modulation of prefrontal-parietal-striatal brain systems during arithmetic and temporal transformations in working memory. J Neurosci. 2007 Dec 5; 27(49): 13393–401. [PubMed: 18057197]

15. Cohen JR, Gallen CL, Jacobs EG, Lee TG, D'Esposito M. Quantifying the reconfiguration of intrinsic networks during working memory. PLoS One. 2014 Jan.9(9):e106636. [PubMed: 25191704]

16. McNab, F., Klingberg, T. Nat Neurosci. Vol. 11. Nature Publishing Group; 2008 Jan. Prefrontal cortex and basal ganglia control access to working memory; p. 103-7.

17. Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci. 1996 Mar 1; 16(5):1936–47. [PubMed: 8774460]

18. Collins AGE, Frank MJ. Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive.

19. Frank MJ, Seeberger LC, O'reilly RC. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science. 2004 Dec 10; 306(5703):1940–3. [PubMed: 15528409]

20. Pessiglione M, Petrovic P, Daunizeau J, Palminteri S, Dolan RJ, Frith CD. Subliminal instrumental conditioning demonstrated in the human brain. Neuron. 2008 Aug 28; 59(4):561–7. [PubMed: 18760693]

21. Frank MJ, Moustafa Aa, Haughey HM, Curran T, Hutchison KE. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl Acad Sci U S A. 2007 Oct 9; 104(41):16311–6. [PubMed: 17913879]

22. Deserno L, Boehme R, Heinz A, Schlagenhauf F. Reinforcement Learning and Dopamine in Schizophrenia: Dimensions of Symptoms or Specific Features of a Disease Group? Front psychiatry. 2013 Jan.4(December):172. [PubMed: 24391603]

23. Gold, JM., Strauss, GP., Waltz, Ja, Robinson, BM., Brown, JK., Frank, MJ. Biol Psychiatry. Elsevier; 2013 Feb 7. Negative Symptoms of Schizophrenia Are Associated with Abnormal Effort-Cost Computations; p. 1-7.

24. Kool, W., Botvinick, M. J Exp Psychol Gen. Vol. 143. American Psychological Association; 2014. A labor/leisure tradeoff in cognitive control; p. 131-41.

25. Westbrook, A., Braver, TS. Cogn Affect Behav Neurosci. Vol. 15. Springer US; 2015 Jun 12. Cognitive effort: A neuroeconomic approach; p. 395-415.

26. Shenhav, A., Botvinick, MM., Cohen, JD. Neuron. Vol. 79. Elsevier Inc.; 2013 Jul 24. The expected value of control: an integrative theory of anterior cingulate cortex function; p. 217-40.

27. Cavanagh, JF., Masters, SE., Bath, K., Frank, MJ. Nat Commun. Vol. 5. Nature Publishing Group; 2014. Conflict acts as an implicit cost in reinforcement learning; p. 5394

28. Cowan N. The Magical Mystery Four: How is Working Memory Capacity Limited, and Why? Curr Dir Psychol Sci a J Am Psychol Soc. 2010 Feb 1; 19(1):51–7.

29. Doll BB, Hutchison KE, Frank MJ. Dopaminergic genes predict individual differences in susceptibility to confirmation bias. J Neurosci. 2011 Apr 20; 31(16):6188–98. [PubMed: 21508242]

30. Cockburn, J., Collins, AGE., Frank, MJ. Neuron. Elsevier Inc.; 2014 Jul. A Reinforcement Learning Mechanism Responsible for the Valuation of Free Choice; p. 1-7.

31. Cox, SML., Frank, MJ., Larcher, K., Fellows, LK., Clark, Ca, Leyton, M., et al. Neuroimage. Elsevier B.V.; 2015 Jan 3. Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes.

32. Dreisbach G, Fischer R. Conflicts as aversive signals. Brain Cogn. 2012; 78(2):94–8. [PubMed: 22218295]

33. Fritz, J., Dreisbach, G. Cogn Affect Behav Neurosci. Vol. 13. Springer-Verlag; 2013 Jun 10. Conflicts as aversive signals: Conflict priming increases negative judgments for neutral stimuli; p. 311-7.

34. Gold, JM., Hahn, B., Strauss, GP., Waltz, JA. Neuropsychol Rev. Vol. 19. Springer US; 2009 Sep 19. Turning it Upside Down: Areas of Preserved Cognitive Function in Schizophrenia; p. 294-311.

35. Heerey EA, Bell-Warren KR, Gold JM. Decision-Making Impairments in the Context of Intact Reward Sensitivity in Schizophrenia. Biol Psychiatry. 2008; 64(1):62–9. [PubMed: 18377874]

36. Waltz, Ja, Kasanova, Z., Ross, TJ., Salmeron, BJ., McMahon, RP., Gold, JM., et al. The roles of reward, default, and executive control networks in set-shifting impairments in schizophrenia. PLoS One. 2013 Jan.8(2):e57257. [PubMed: 23468948]

37. Schlagenhauf, F., Huys, QJM., Deserno, L., Rapp, Ma, Beck, A., Heinze, HJ., et al. Neuroimage. Elsevier B.V.; 2013 Nov 27. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients.

38. Dowd, EC., Frank, MJ., Collins, A., Gold, JM., Barch, DM. Biol Psychiatry Cogn Neurosci Neuroimaging. Elsevier; 2016. Archival Report Probabilistic Reinforcement Learning in Patients With Schizophrenia : Relationships to Anhedonia and Avolition; p. 1-14.

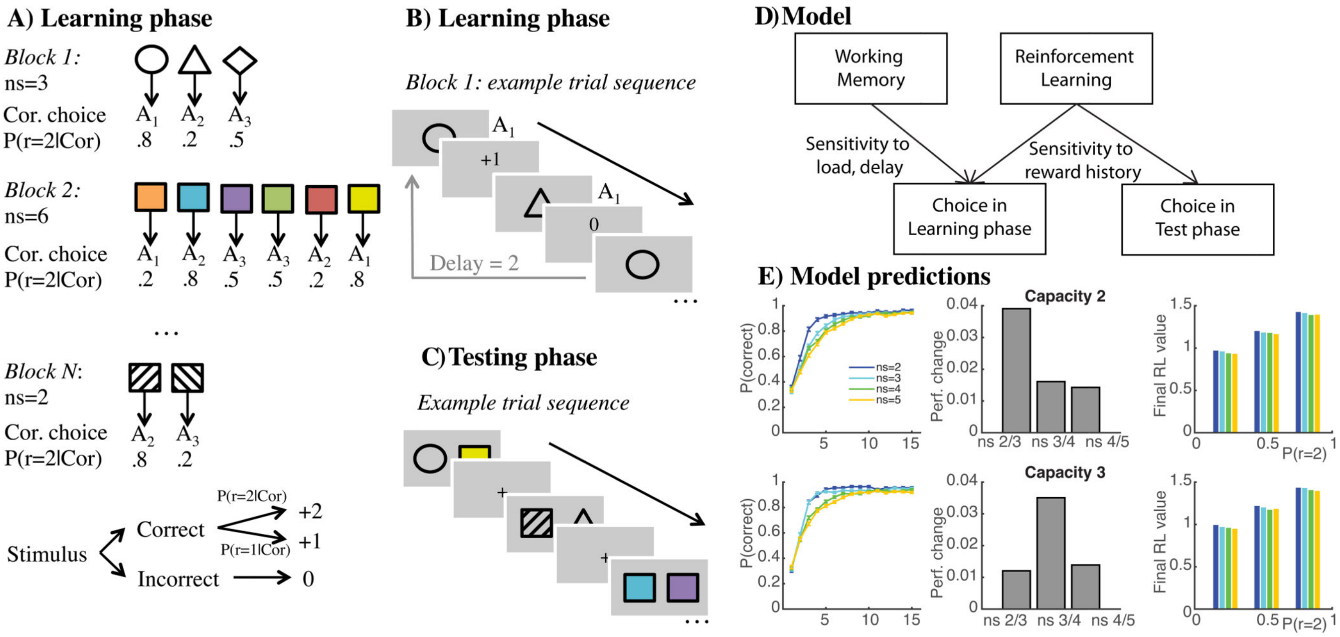

Figure 1. Experimental protocol

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 vs. 1 points 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) In 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 and reinforcement learning), making behavior sensitive to load, delay, and reward history. In contrast, test performance is only dependent on RL, such that if RL and WM are independent, choice should only depend on reward history. E) 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: Learning curves indicate the proportion of correct trials as a function of the number of encounters with given stimuli in different set sizes. Middle: 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 for capacity 3. Right: assuming RL 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 vs. 2 points in correct trials.

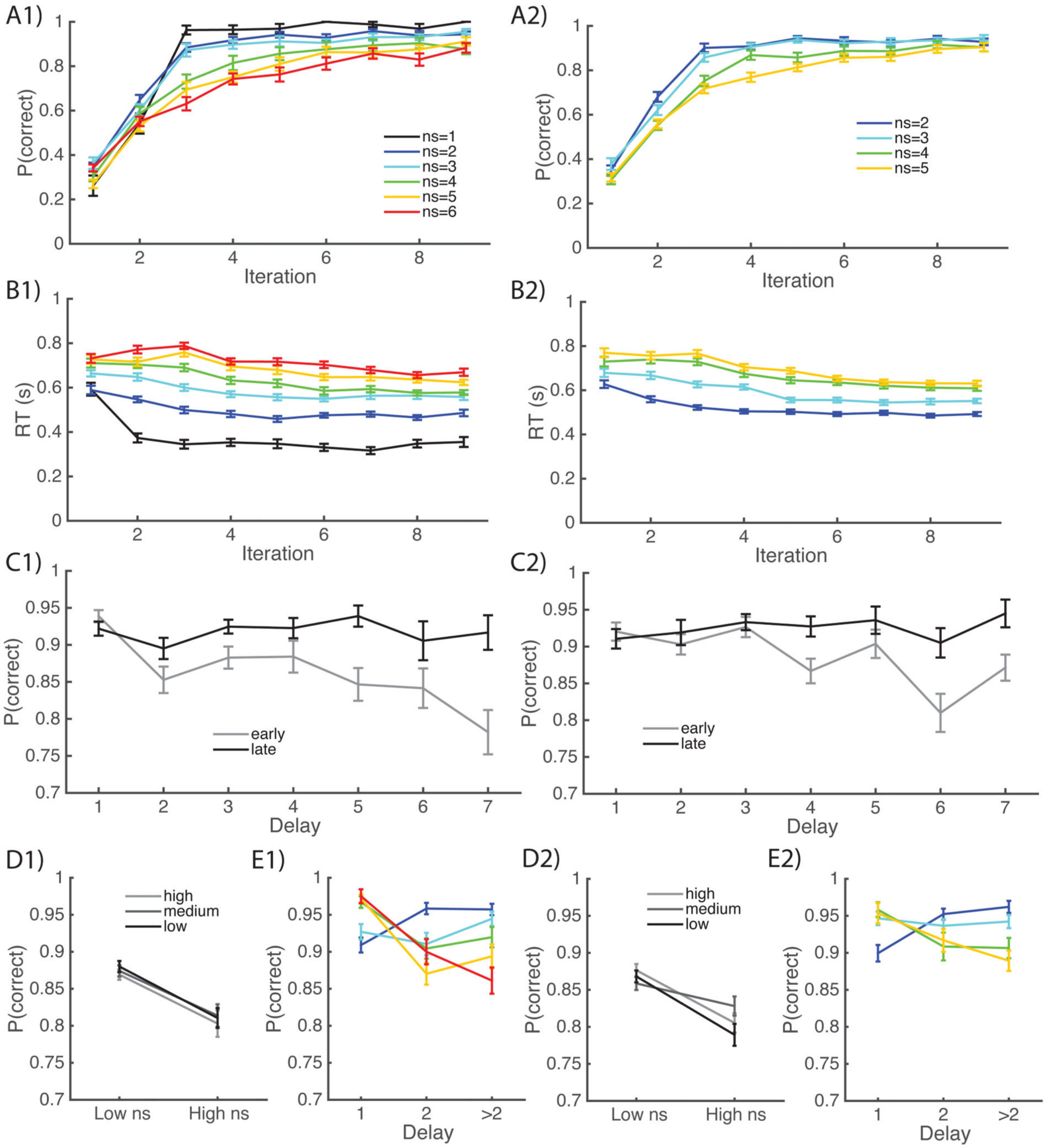

Figure 2.

A-B) learning curves show the proportion of correct trials and mean reaction times as a function of the encounter number of each stimulus, for different set sizes (ns). Left/right columns show results from experiment 1, 2. 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 vs. 1 when correct choice is made.

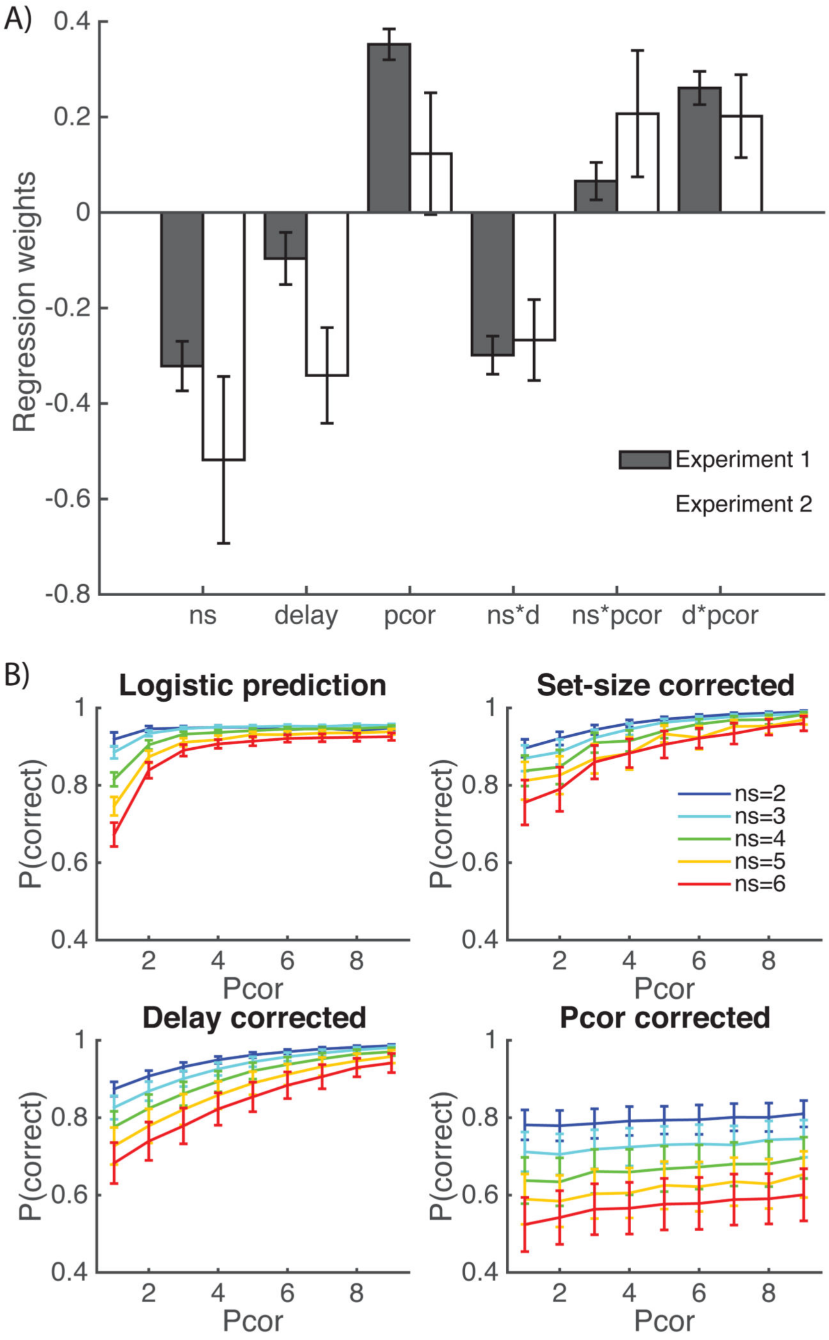

Figure 3.

Learning phase: A) Results from the logistic show consistent effects of set size, delay, number of previous corrects (Pcor), and interactions (except set size with pcor). Error bars indicate standard error of the mean. B) Experiment 1 Logistic regression predictions (top left) 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.

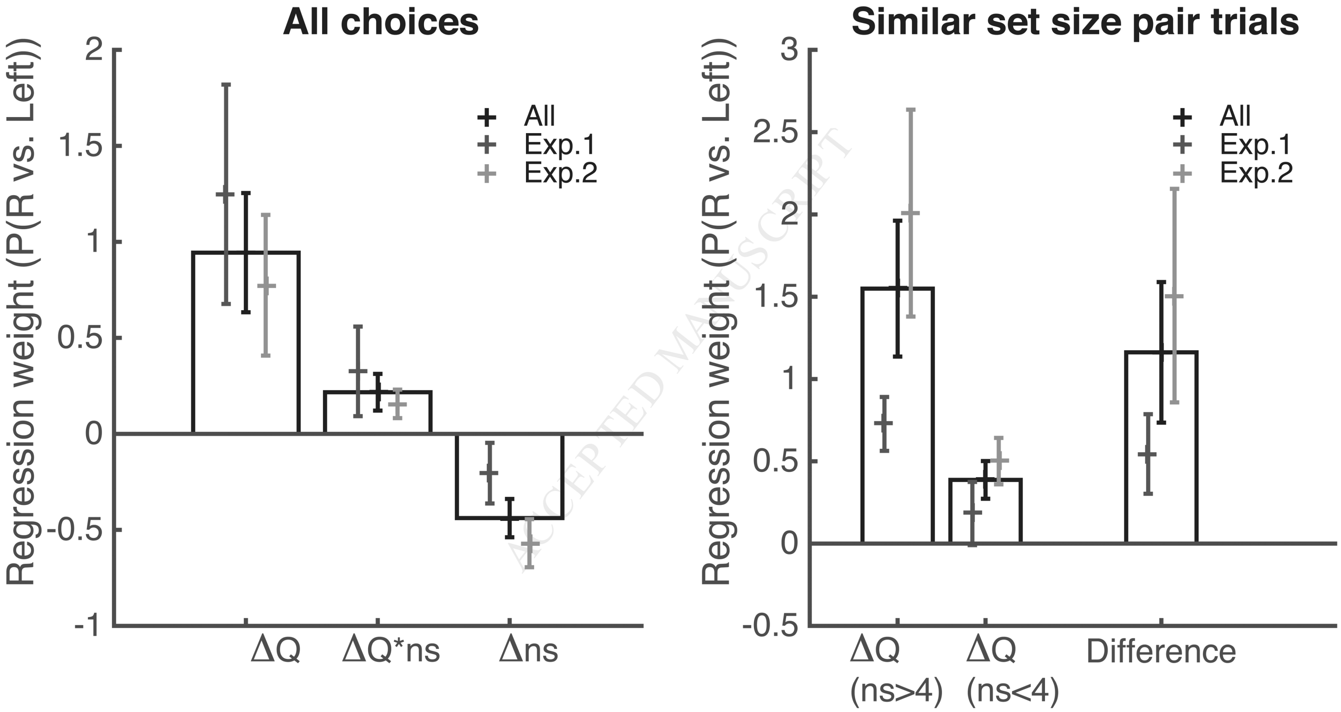

Figure 4.

Test phase results. A) We analyze choice of the right vs. left image in the test phase as a function of the value difference *ΔQ=value(right image) – value(left image)*, the set size difference *Δns*, *ΔQ*ns* the interaction of the mean set size of the two images with the value difference, as well as other regressors of non interest. We find a significant effect of all three factors across both experiments. B) the effect of value difference is significantly stronger in high set sizes than in low set sizes, indicating that RL was more efficient under high load, thus highlighting an interaction of RL with WM.

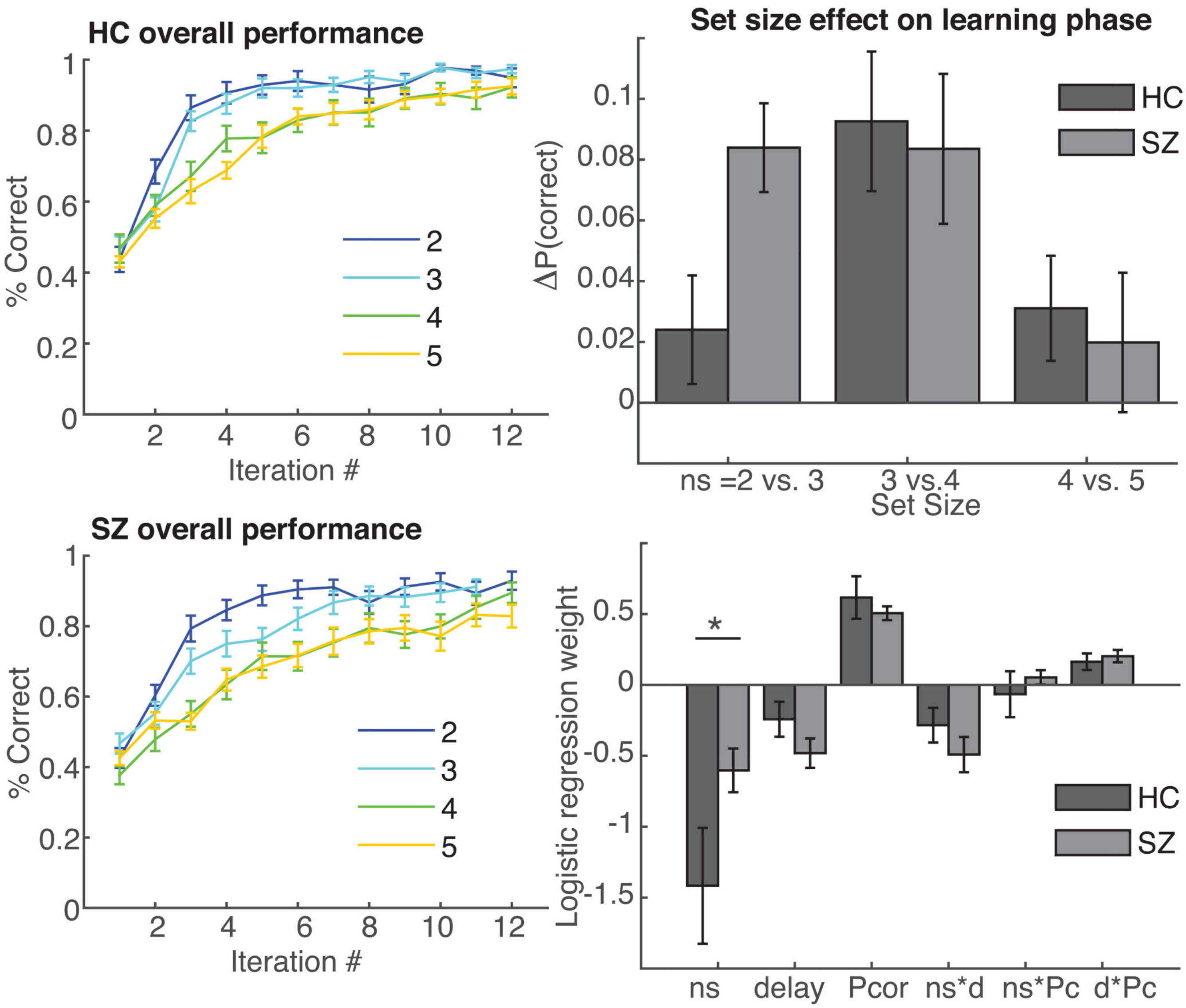

Figure 5.

Schizophrenia learning phase results replicate our previous finding that WM contributes to learning impairment. *Left,* learning curves (see Fig. 2) show slower learning for people with schizophrenia than control*. Top right:* change in performance from set size 2 to 3 is significantly higher in people with schizophrenia than controls. HC pattern matches a capacity 3 model simulation (Fig 1 E), while PSZ pattern matches a mixture of capacity 2 and capacity 3 model simulation. *Bottom right*: logistic regression analysis shows only a difference in the set size effect between groups, implicating the working memory mechanism.

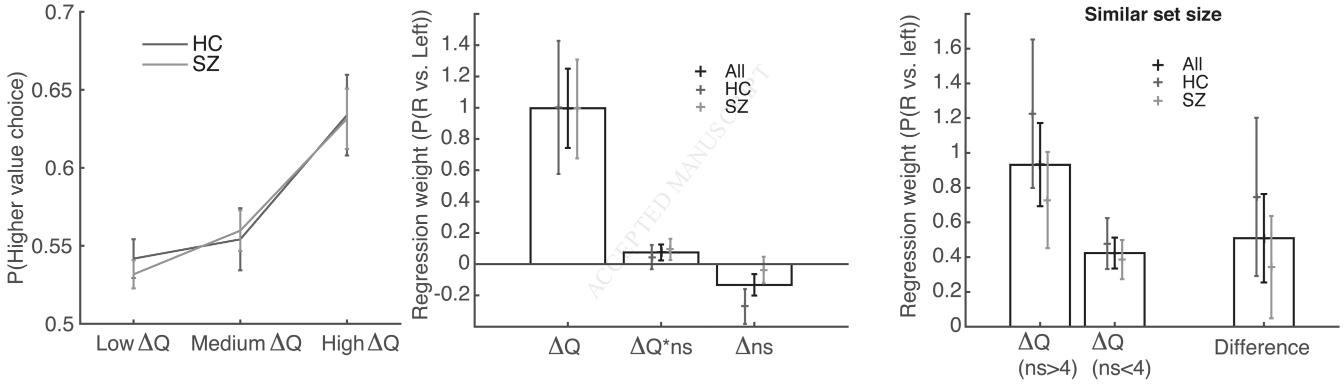

Figure 6.

Schizophrenia test phase supports our prediction that RL-dependent value learning is unimpaired in people with schizophrenia. Left: 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 PSZ and controls (HC). Middle: logistic regression analysis of the test phase supported our finding supported our finding that PSZ and controls were equally sensitive to value difference. We found an effort effect in HC, but not in PSZ. Right: both groups were more sensitive to value difference in high than low set sizes, supporting our previous result.

Table 1

Experiment 3 Demographics

	HC	PSZ[a]	P value
n	32	46	
Age (years): mean (SD)	37.14 (10.21)	37.81 (8.97)	0.76
Education (years): mean (SD)			
Participant	15.06 (2.12)	13.27 (2.37)	0.001
Maternal	13.75 (2.21)	14.02 (2.93)	0.66
Paternal	14.20 (3.68)	13.74 (3.52)	0.60
Sex, Male/Female, *n*	21/11	28/18	0.58
Race/ethnicity, *n*			0.85
African American	12	18	
White	17	26	
Other	3	2	

[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.