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 Biological Psychiatry September 15, 2017; 82:431–439 www.sobp.org/journal

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

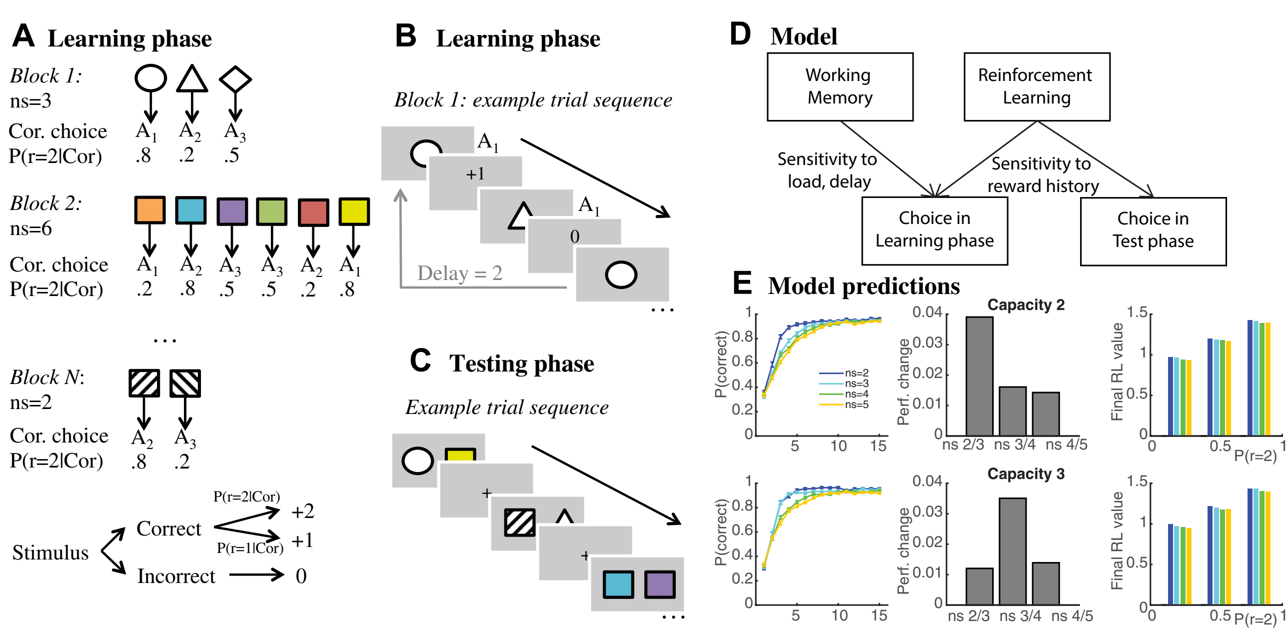

Figure 1. Experimental protocol. **(A)** Learning phase. Participants learn to select one of three actions (key presses $A_{1=3}$) for each stimulus in a block using reward feedback. Incorrect choices lead to feedback 0, while correct choices lead to reward, either +1 or +2 points, probabilistically. 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.