Dual Effects of Dual-Tasking on Instrumental Learning

Huang Ham[3], Samuel D. McDougle[4], Anne G.E. Collins[1,2]

Department of Psychology, University of California, Berkeley[1]

Helen Wills Neuroscience Institute, University of California, Berkeley[2]

Department of Psychology, Princeton University[3]

Department of Psychology, Yale University[4]

Abstract

How automatic is reinforcement learning (RL)? Here, using a recent computational framework that separates contributions from working memory versus RL during instrumental learning, we asked if taxing higher executive functions influences a putatively lower-level, procedural RL system. Across three experiments, we found that dual-tasking could indeed disrupt RL, even when isolating RL from working memory contributions to behavior. These results speak to methodological considerations in the use of dual tasks during learning, suggesting that cognitive load can interfere with multiple learning and memory systems simultaneously. Moreover, our results point to a less constrained conception of RL as a putatively low-level procedural system, supporting a view that tight links exist between executive function and subcortical learning processes.

Keywords: Reinforcement learning; Working memory; Executive control; Computational modeling; Dual-Tasks

Dual Effects of Dual-Tasking on Instrumental Learning

Introduction

The study of instrumental learning (learning to select actions that lead to rewards) typically focuses on the *reinforcement learning* process (RL), which is well captured by a computational framework that formalizes reward as a teaching signal to estimate expected values (Rescorla, 1972; Sutton & Barto, 1998). Although RL is a powerful learning system, human beings also utilize higher-level executive functions during instrumental learning tasks, such as working memory (WM) and attention. A growing body of research suggests that executive functions like working memory and attention shape the learning of simple instrumental policies alongside reinforcement learning (A. G. Collins & Frank, 2012; Leong, Radulescu, Daniel, DeWoskin & Niv, 2017; Rmus, McDougle & Collins, 2021; A. Yoo & Collins, 2022). Executive functions typically require top-down cognitive control, process information explicitly, and operate on a shorter time span, whereas reinforcement learning operates more implicitly, and over a longer time span (A. G. Collins, 2018). For example, executive functions could aid instrumental learning by directing attention to relevant reward signals and contextual cues, and encode these sources of information explicitly in working memory (such as explicitly remembering that one action yielded a reward but another action did not). Due to the intrinsic capacity limitations of working memory, however, people are unlikely to be able to explicitly remember sufficient information about reward-action contingencies over longer periods of time. Nevertheless, even without explicit memory, people are still able to implicitly learn to choose more rewarding actions over less rewarding ones (Cortese, Lau & Kawato, 2020; Gabrieli, 1998; Pessiglione et al., 2008; Shohamy, 2011; Wilkinson & Jahanshahi, 2007), as demonstrated, for example, by their ability to learn more information than can be held in working memory (A. G. Collins & Frank, 2012). This phenomenon is typically attributed to the reinforcement learning (RL) process.

Across various populations, studies have shown that working memory and reinforcement learning indeed operate in parallel during simple instrumental learning tasks, and compete for action control (A. G. Collins & Frank, 2012; Master et al., 2020; Viejo, Khamassi, Brovelli & Girard, 2015). These findings can be formalized in computational models that include both RL and WM - such models are designed to capture human behavioral and neural data in simple instrumental learning contexts (A. G. Collins, Ciullo, Frank & Badre, 2017; A. G. Collins & Frank, 2018; Viejo et al., 2015).

While it is clear that both WM and RL can contribute to human reward learning, what is poorly understood, however, is whether reinforcement learning processes are *functionally independent* of executive functions, or if the two systems *interact* with each other. Past research has typically framed RL as a closed-loop, lower-level process that does not strongly rely on higher-level cognitive inputs. That is, RL is often thought of as being a procedural learning system. However, recent research has challenged this view by suggesting multiple ways in which RL computations appear to be tightly linked to executive functions, including attention (Leong et al., 2017; Niv et al., 2015) , abstract motivational goals (McDougle, Ballard, Baribault, Bishop & Collins, 2021; Sinclair, Wang & Adcock, 2023), and working memory (A. Collins, Ciullo, Frank & Badre, 2017; A. G. Collins, 2018; A. G. Collins & Frank, 2018; Rmus et al., 2021; A. Yoo & Collins, 2022). To our knowledge, minimal prior work has applied causal experimental tests on links between executive functions and reinforcement learning processes that perturb executive function while also measuring its direct contributions to learning. Without doing so, it is difficult to know if perturbing an executive function (e.g., WM) during learning simply disrupts that specific function's contributions to behavior, or if 'downstream' effects on the RL system are also induced. If executive functions contribute to instrumental learning independently, taxing them would *not* impact the reinforcement learning process, and indeed only impact learning behavior through executive function contributions themselves. On the other hand, if RL is not fully separable from parallel executive function contributions to learning, perturbing executive functions should additionally impact the reinforcement learning process. This impact on RL could be either facilitating (leading to faster learning of rewarding actions) or inhibitory (leading to slower learning).

In three experiments, we tested these hypotheses by directly perturbing executive functions using a classic *"dual-task"* manipulation during an instrumental learning paradigm that is optimized to disentangle RL from WM. Dual-tasks are a common procedure for taxing executive function and have been deployed across a range of cognitive and learning tasks (Baddeley, 1992; D'Esposito et al., 1995; Economides, Kurth-Nelson, Lübbert, Guitart-Masip & Dolan, 2015). We designed two "dual-task" conditions which only differed in when the dual task occurred within the flow of the experiment: "Task-Overlap" and "Task-Switch". The "Task-Overlap" condition directly taxed executive function by presenting extra information for the participant to remember while simultaneously performing the learning task. The "Task-Switch" condition freed participants from any extra working memory load during the choice and feedback process, but required them to engage in the recruitment of executive functions between learning trials. We performed 3 experiments: In the first 2, we compared the (standard) Single-Task condition with the "Task-Overlap" condition, and varied the single-task inter-trial interval across experiments to control for timing differences between single- and dual-task settings (see Methods). In the third experiment, we compared the "Task-Overlap" condition and the "Task-Switch" condition to each other.

Our overarching goal was to use a computational modeling framework (the "RLWM" model) that captures reward learning behavior with separable WM and RL modules (A. G. Collins & Frank, 2012), allowing us to examine how taxing executive function through a dual task might affect different sub-components of instrumental learning. The "RLWM" model was crucial for testing the effect of perturbing executive functions on reinforcement learning, as behavioral data alone (such as average accuracy metrics) can depend on both mechanisms. All experiment materials, data, and analysis code are publicly available at https://osf.io/zutka/?view_only=022f6bc1c9324df790eabe24e200286d.

Methods

Participants

Participants in all three experiments (N1 = 31, N2 = 31, N3 = 33) were recruited through the University of California Berkeley's SONA platform and earned class credit for their participation. In experiment 1, 21 females and 10 males participated with a mean age of 20.47. In experiment 2, 17 females and 14 males participated with a mean age of 21.32. In experiment 3, 26 females and 7 males participated with a mean age of 21.43. No participants were excluded. The experimental protocol was approved by the university's local ethics committee. Written, informed consent was obtained from all participants prior to their participation.

Experimental Procedure

Experiment 1. Participants were seated in front of a computer monitor and had their hands comfortably positioned on a computer keyboard. They then proceeded to the main experiment which was a computerized task written using Psychtoolbox (version 3.0.10) on Matlab (version R2016a). The main goal for the participants was to learn which key (out of 3 candidate keys) on the keyboard was associated with each stimulus presented on the screen. We used images from (A. G. Collins et al., 2017) as stimuli in our task.

After instruction and practice (aimed to familiarize the participant with the task), the task had *two phases*: learning, and testing. In the learning phase, participants attempted to learn multiple stimulus-response pairs in separate, independent blocks. In the testing phase, all stimuli from all learning phase blocks were displayed again in a random sequence, and participants responded but did not receive correct/incorrect feedback, allowing us to probe long-term retention of learned information, independent of WM.

The learning phase (figure 1) consisted of 10 independent blocks of trials, but the last block only served as a buffer between the learning and the testing phase and thus was excluded from later analyses. In each trial, participants saw an image presented on the screen and pressed one of the three keys in response. A block consisted of either 2, 3, or 6 image-key associations to learn and 12 iterations per image, pseudo-randomly interleaved to control for an approximately uniform distribution of delays between iterations of the same stimulus. Each block used a separate set of images to be learned, consisting of easily distinguished and named examplars of a category (e.g. vegetables, farm animals, etc. Aspen H Yoo, Keglovits and Collins, 2023). At the beginning of each block, participants saw all the images that they would encounter in that block for familiarization. Across blocks, the *set size* of the instrumental learning task was varied among 2, 3, and 6 (A. G. Collins & Frank, 2012). That is, in each block participants had to either learn 2, 3, or 6 stimulus-response associations, a manipulation that is critical to delineating WM and RL in our modeling framework (A. G. Collins & Frank, 2012). Stimuli were never repeated across blocks. The learning phase also included two conditions: *Dual-Task* and *Single-Task*, across blocks. In the Dual-Task condition, two blocks were performed at each set size, and in the Single-Task condition, one block was performed each at set sizes 3 and 6, and two blocks at set size 2. The block order was pseudo-randomized except the last (10th) block. The last block, which was used as a buffer, always had set size 2 and trials in the Single-Task condition.

In the Dual-Task condition, a secondary task — the *number judgment task* — was performed in addition to the instrumental learning task (Economides et al., 2015). For this task, two numbers were simultaneously displayed side-by-side with varying font sizes and integer values (e.g., a large font “2” on the left and a smaller font “6” on the right). Participants were asked to make either a “size” or “value” judgment of the number stimuli by pressing a key that corresponded to the position of either the visually larger number (e.g., “2”, or left button) or the higher-value number (e.g., “6”, right button; figure 1) . The particular judgment required (value versus size) was randomly selected on each trial. Approximately 80% of trials consisted of conflict trials, where the visually larger integer was smaller in value and vice-versa. The specific two integers presented were drawn randomly from $[0, 9]$ without replacement.

In Dual-Task blocks, the trial structure was as follows: Participants viewed one of the learning stimuli on the screen and two numbers positioned above the stimulus (Figure 1). The numbers were displayed for 0.3 seconds. The learning stimulus was continually displayed either until the participant responded with one of the three possible actions ("j", "k", or "l" with their right index, middle, or ring finger), or if 1.5 seconds had elapsed. If the response designated as correct for that stimulus was made, +1 "points" were displayed on the screen. If an incorrect response was given, 0 points were displayed. If the reaction time exceeded 1.5 seconds, the message "please respond faster" was displayed, and if the response was faster than 0.15 seconds the message "too fast" was displayed. The feedback to the instrumental learning task was displayed for 1 second. Critically, after receiving feedback for the instrumental learning task, the participant was then asked to make either a "size" or "value" judgment of the previously-displayed numbers ("a" or "d" with their left ring and index fingers, corresponding to the number displayed on the left or right, respectively). Participants had up to 1 second to respond to the number judgment task, but if they responded in less than 1 second, they would still need to wait until the end of the second before seeing feedback. Feedback was then given for the number judgment task ("correct", "incorrect", "please respond faster", or "too fast") and was displayed for 1 second as well. An inter-trial interval of 1.5 seconds (minus the reaction time of the instrumental learning task) then occurred, which consisted of a white fixation cross displayed in the center of the screen. The interval was computed as such to control for the total trial duration. Therefore the total trial duration was 4.5 seconds.

In Single-Task blocks, participants didn't need to perform the number judgment task, but only needed to perform the instrumental learning task. Therefore, there was no number displayed above the learning stimulus and there was no question about the numbers following the feedback for the instrumental learning task. To ensure that the total trial length was the same as in the Dual-Task condition, the inter-trial interval was 3.5 seconds minus the reaction time.

To become familiarized with the tasks, participants performed the practice phase with three unique practice rounds before the learning blocks began: They first practiced the instrumental learning task on its own (10 trials), followed by the "number judgment" secondary task on its own (10 trials), then the Dual-task condition (10 trials). Experimenter instructions emphasized that participants should focus on performing equally well on both tasks in all blocks.

After the learning phase, participants proceeded to perform a surprise testing phase. In the testing phase, the screen first displayed the instruction telling them that they would see images that they had encountered previously and that they needed to respond by retrieving the action that they originally learned was correct for that image (j, k, or l key). Similar to the learning phase, participants' response to a trial was valid if made between 0.15 and 1.5 seconds from the onset of the image. Unlike in the learning phase, however, no feedback followed their actions and there was no inter-trial interval. The testing phase was not divided into blocks, and all the images in the learning block were shuffled and presented in sequence at the center of the screen. Each image appeared four times in total in this shuffled sequence. The testing phase was included to provide a measure of long-term associations formed through RL, without immediate contributions from working memory processes (contrary to the learning phase where information was available within a short time frame). Because the information encoded in the RL system is retained for a longer period of time than the information encoded in working memory, we can attribute participants' performance in the testing phase more to the learning outcome of the RL system (A. G. Collins, 2018).

Experiment 2. While experiment 1 controlled for the total trial duration between the Single-Task and the Dual-Task condition, the inter-trial intervals in the Single-Task condition were substantially longer than in the Dual-Task condition, potentially introducing a confound. In experiment 2, we instead controlled for the inter-trial interval between the two conditions. Experiment 2 (figure 1) was identical to experiment 1 except that the inter-trial interval in the Single-Task condition was the same as the inter-trial interval in the Dual-Task condition, which was 1.5 seconds minus the reaction time. Therefore, unlike in Experiment 1, where the total trial duration was the same between the two conditions, the trial duration of the Single-Task condition was shorter than the trial duration of the Dual-Task condition in Experiment 2.

Experiment 3. While the previous 2 experiments controlled for the differences in inter-trial interval and trial duration, they could not identify whether the potential Dual-Task effect comes from simply having to switch tasks during learning, or from having to hold two numbers in memory while making decisions. To disentangle these two possibilities, we designed experiment 3 (figure 1).

The learning phase of experiment 3 did not have Single-Task conditions, but instead, it consisted of 2 different Dual-Task conditions: *Task-Overlap* and *Task-Switch*. The Task-Overlap condition is exactly the same as the Dual-Task condition in experiments 1 and 2. Thus, in Task-Overlap blocks, the number task and instrumental task were performed simultaneously – the number sizes and values had to be encoded and maintained while the correct stimulus-response association was being learned and/or retrieved. In contrast, in Task-Switch blocks, the same two tasks were performed but in succession – a complete trial of the instrumental learning task was performed (learning stimulus, response, feedback), followed by a complete trial of the number judgment task (number stimuli, response, feedback). In the instrumental learning task, same as the Single-Task trials in experiments 1 and 2, participants viewed one of the learning stimuli on the screen **without** the additional two numbers above them. The learning stimulus was continually displayed either until the participant responded with a valid keypress, or if 1.5 seconds had elapsed. The feedback was then displayed for 1 second. After having received feedback for the instrumental learning task, the participant was then asked to make either a “size” or “value” judgment of two numbers. Unlike in the Task-Overlap condition, the two numbers were displayed right above the question, so participants could make the judgment while looking at the two numbers. Participants had also up to 1 second to respond to the number judgment task, but if they responded less than 1 second, they would still need to wait until the end of the second before seeing feedback. The feedback was displayed for 1 second as well. The feedback mechanism for both the instrumental learning task and the number task is the same as in experiments 1 and 2. Afterward, an inter-trial interval of 1.5

seconds (minus the reaction time of the instrumental learning task) occurred. The interval was computed as such to control for equal total trial duration (4.5s) across the two conditions.

In sum, the only difference between the Task-Switch and the Task-Overlap condition was that in the Task-Overlap condition, the two numbers appeared simultaneously with the instrumental task stimulus for 0.3 seconds, and thus participants needed to hold these numbers in memory during the instrumental learning task, but in the Task-Switch condition participants did not need to hold them in memory while performing the instrumental learning task. That is, the Task-Switch condition was included to benchmark the global effects of taxing executive function without requiring secondary task representations to occupy working memory during the choice and feedback phases of the instrumental task.

All other aspects of the experiment were largely the same as experiments 1 and 2, replacing the Single-Task condition with the Task-Switch condition and replacing the Dual-Task condition with the Task-Overlap condition. Particularly, in the Task-Overlap condition, two blocks were performed at each set size, and in the Task-Switch condition, one block was performed each at set sizes 3 and 6, and two blocks at set size 2. The last (10th) block always had a set size of 2 and trials in the Task-Switch condition, i.e., the easiest type of block, serving as a buffer between the learning and the testing phase and thus was excluded from all analyses. In the practice phase, participants performed 10 more trials of Task-Switch tasks after the 10 trials of the practice Task-Overlap trials.

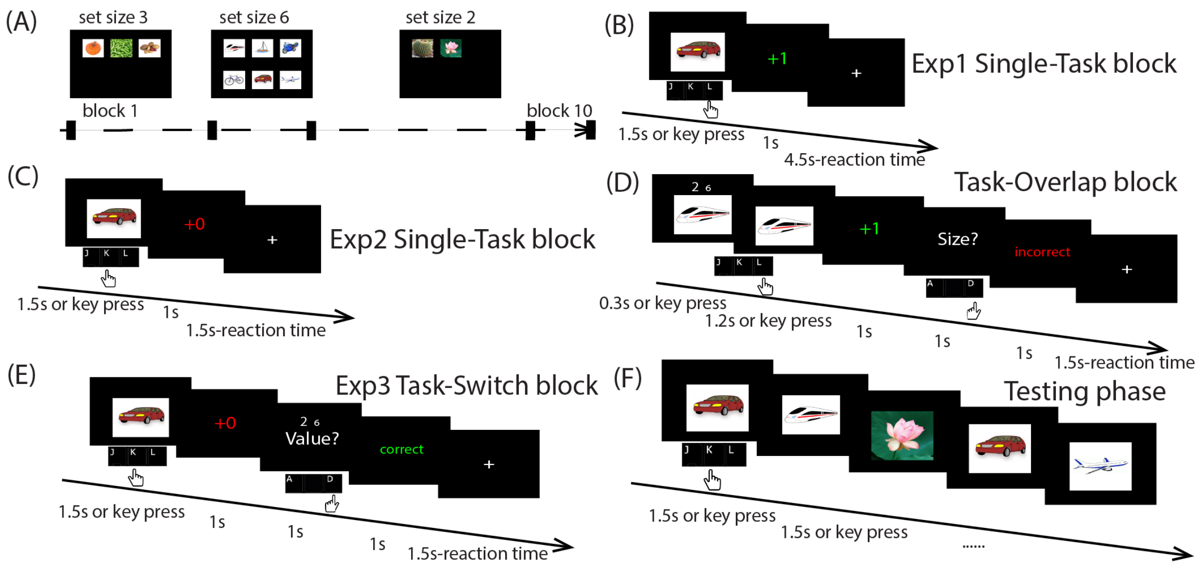

Figure 1. Task Design: (A) Block structure of the learning phase (all experiments): Participants performed 10 independent blocks of the instrumental learning task. The 10th block served only as a buffer between the learning and testing phase. Thus it was removed from all analyses. Participants saw a display of all possible stimuli in the block at the beginning of each block. (B, C) Single-Task blocks (experiment 1 and 2): regular instrumental learning task, each controlling for the total trial duration (B) or the inter-trial interval (C). (D) Main dual-task manipulation: Task-Overlap blocks (all experiments): participants had to remember the two numbers presented concurrently with the stimulus. After making a stimulus-dependent key-press (e.g. here L), and obtaining feedback (here a correct +1), participants were asked to perform a size or value judgment based on the remembered numbers. (E) Task-Switch blocks (experiment 3): the two numbers for the secondary task were presented after participants received the trial's feedback, such that participants did not have to remember the two numbers but only needed to judge the numbers between learning trials. (F) Testing phase (all experiments): Each image repeated four times at randomized places in the sequence. No feedback was given.

Statistical analyses

All statistical analyses were done in R (version 4.3.1). To calculate the standard error of the mean, we used the *std.error* function in the *plotrix* package (version 3.8.2). To perform the Wilcoxon test, we used the *wilcox_test* function in the *rstatix* package (version 0.7.2). To statistically quantify the impact of different task variables on performance, we performed a two-way ANOVA and a *mixed-effect regression* analysis. To perform two-way ANOVA, we used the *aov* function. The dependent variable was average accuracy. The independent variables included were set size (3 levels: 2, 3, 6) and dual-task condition (2 levels: Task-overlap vs the control condition depending on the experiment) and the interaction term. The mean and standard deviation are reported in supplemental table 1. To perform *mixed-effect regression* analysis, we used the *mixed* function in package *afex* (version 1.3.0), with model comparison method set to *LRT*, representing a likelihood ratio test. All continuous variables were scaled before passing them into the regression. We set the correct/incorrect responses as the outcome variable and subject identification number as the random intercept. For the learning phase data, we passed in four task variables as predictors: *condition, set size, delay (i.e., the number of intervening trials between the current and previous viewings of a specific stimulus), and cumulative reward (i.e., the number of successful trials with the current stimulus).* For the testing phase data, we passed in three task variables as predictors of performance: *condition*, *set size*, and *asymptotic learning phase performance.* We obtained the condition and set size of the stimuli presented in the testing phase by referring to the condition and set size those stimuli had belonged to during the preceding learning phase. The asymptotic rate of performance for each stimulus was obtained by computing the average correctness of the last 3 trials for that stimulus from the learning phase.

The RLWM Computational Model

Here we present the details of the "RLWM" model architecture, which functions as the basic foundation of our model-dependent analyses (A. G. Collins & Frank, 2012). The model was designed to fit participants' choices in this instrumental learning task, and capture simultaneous contributions from working memory and reinforcement learning. Prior work showed that this model outperforms alternative models that do not include a hybrid RL + WM structure; indeed, other models could not capture the patterns of behavior that reveal the dissociable contributions of RL and WM on the performance in this instrumental learning task, in particular the strong effects of set size on accuracy (A. G. Collins, 2018; A. G. Collins & Frank, 2012; Rac-Lubashevsky, Cremer, Collins, Frank & Schwabe, 2023; Rmus et al., 2023). Therefore we rely on the RLWM computational framework to further examine the separate effects of perturbing executive functions on the RL and the WM system.

The RLWM model models the learning of stimulus-action values using a variant of a typical reinforcement learning model (Sutton & Barto, 1998). The model relies on two main variables representing the task environment. The first one is the state $s \in S$ where S represents the full stimulus/state space within a block (i.e., all the possible images that could appear). In our experiment, $|S| \in \{2, 3, 6\}$. The second variable is the action $a \in A$ where A is the full action space (i.e., j, k, l). In our experiment, $|A| = 3$ because there were three possible buttons to press as a response to the instrumental learning task. The algorithm proceeds in two stages, as introduced in the introduction: the value updating stage and the policy formation stage. In the value updating stage, for stimulus s and action a on trial t, the model estimates an expected value (i.e., the Q value) $\boldsymbol{Q}(s_t, a_t)$ by performing an update using the delta rule (equation 2; Rescorla, 1972):

$$\boldsymbol{Q}_{t+1}(s_t, a_t) = \boldsymbol{Q}_t(s_t, a_t) + \alpha\delta_t \tag{1}$$

$$\delta_t = r_t - \boldsymbol{Q}_t(s_t, a_t) \tag{2}$$

where α represents the **learning rate** and $\boldsymbol{Q}_t$ is a $|S| \times |A|$ matrix encoding all Q values given a trial t. $\boldsymbol{Q}_0$ is initialized as a uniform matrix of $\frac{1}{|A|}$. $\delta \in [0, 1]$ is the reward prediction error, and $r \in \{0, 1\}$ is the (binary) reward received. Critically, the model captures the parallel recruitment of working memory (WM) and reinforcement learning (RL) by training two simultaneous learning modules: The reinforcement learning module is described by equation 1. The working memory module is formally similar but has a learning rate of $\alpha = 1$ (algebraically equivalent to equation 3). Thus, the working memory delta rule has perfect retention of the outcome of the previous trial with stimulus s_t, reflecting rapid learning of stimulus-response pairs that is qualitatively distinct from classic reinforcement learning. Working memory is also vulnerable to forgetting (Posner & Keele, 1967): The model captures trial-by-trial decay of stimulus-action weights W (equation 4),

$$\boldsymbol{W}_t(s_t, a_t) = r_t \tag{3}$$

$$\boldsymbol{W}_{t+1} = \boldsymbol{W}_t + \gamma(\boldsymbol{W}_0 - \boldsymbol{W}_t) \tag{4}$$

where $\gamma \in [0, 1]$ is the **forgetting parameter** that draws all W weights toward their initial values $\boldsymbol{W_0} = \boldsymbol{Q_0}$. The model also captures a positive learning bias (i.e., the neglect of negative feedback) upon negative prediction errors (i.e., $\delta < 0$). The learning rate α is reduced multiplicatively: $\alpha^- * \alpha$ where $\alpha^- \in [0, 1]$ controls the **learning bias** (higher values cause less bias toward positive feedback, and lower values cause more). Learning bias occurs for *both the reinforcement learning and working memory modules*; in the latter case, the perfect learning rate of 1 is also scaled by α^-.

In the policy formation stage, Q-values and W weights are transformed by the *Softmax function* into a policy, i.e., a vector of probabilities of taking each action. Separate working memory and reinforcement learning policies (represented by *row vectors* π_t^{WM} and π_t^{RL}) are then combined in the calculation of the final policy via a weighted sum (equation 7),

$$\pi_t^{RL} = p(A|s_t) = Softmax(\boldsymbol{Q}(s_t), \beta) = \frac{e^{\beta \boldsymbol{Q}(s_t)}}{\sum_{a \in A} e^{\beta \boldsymbol{Q}(s_t, a)}} \tag{5}$$

$$\pi_t^{WL} = p(A|s_t) = Softmax(\boldsymbol{W}(s_t), \beta) = \frac{e^{\beta \boldsymbol{W}(s_t)}}{\sum_{a \in A} e^{\beta \boldsymbol{W}(s_t, a)}} \tag{6}$$

$$\pi_t = w\pi_t^{WM} + (1 - w)\pi_t^{RL} \tag{7}$$

where $\beta \in [0, \infty)$ represents the inverse softmax temperature and $w \in [0, 1]$ approximates how much working memory contributes to the eventual decision. This value is determined by two free parameters, the **working memory capacity** (i.e., resource limit) $K \in [2, 5]$, and the **initial working memory weight** $\rho \in [0, 1]$,

$$w = \rho * min\Big(1, \frac{K}{|A|}\Big) \tag{8}$$

This equation can be interpreted as the weight given to the working memory module is reduced if the set size exceeds working memory capacity K, in proportion to the ratio of items that can be held in working memory.

Finally, **un-directed decision noise** ($\epsilon \in [0, 1]$) is added to the final weighted policy (π) to capture potential noise during choice (action retrieval),

$$\pi_t \leftarrow \epsilon\Big(\frac{1}{|A|}\Big) + (1 - \epsilon)\pi_t \tag{9}$$

Modeling Procedure

The modeling followed five steps: model fitting, model comparison, parameter recovery, model recovery, and model simulation and validation (Wilson & Collins, 2019). Models were fit to participants' choices using maximum likelihood estimation, by minimizing the negative log likelihood using the MATLAB fmincon function. Parameter constraints were defined as follows: $\alpha, \gamma, \alpha^-, \rho, \epsilon, \in [0, 1]$ and $C \in [2, 5]$. Initial parameter values were randomized within their constraints across fitting iterations. Inverse temperature β was fixed at 100 for all fits and simulations, reflecting optimal parameter recovery results from previous work using this model (Master et al., 2020). Each subject was fit over 100 iterations to avoid local minima in parameter values. Single task and dual task blocks were fit separately to examine the effects of dual-tasking on the fitted model parameters.

Model simulation and validation were performed to ensure that the model's key parameter value correlates with key behavioral features of the data and that the models' learning behavior reproduces a qualitative pattern similar to that of human participants. Model simulations were conducted by simulating the model using each participant's best-fit parameters and their actual observed sequence of stimuli and blocks. Model simulations were performed 100 times per subject and averaged.

Results

Dual-task performance

We first sought to validate the dual-task manipulation by checking that participants performed well in the secondary task. Indeed, participants on average made correct choices in 81.0% of trials of the number task correct ($SE = 0.013$) in the task-overlap condition of experiment 1, 81.0% correct ($SE = 0.016$) in experiment 2, and 82.7% correct ($SE = 0.013$) in experiment 3, well above chance level (50%). Participants also obtained an accuracy of 82.9% ($SE = 0.015$) in the task-switch condition in experiment 3. The number task accuracy of the two conditions in experiment 3 did not significantly differ from each other according to the Wilcoxon test ($U = 559, p = 0.858$). The number task accuracy and reaction time in the Task-overlap condition across all experiments did *not* depend on the congruency (whether the number larger in value was also larger in font size) of the numbers ($U = 636, p = 0.243; U = 459, p = 0.278$). However, in the Task-switch condition in experiment 3, we did observe that participants were more accurate and reacted faster in the congruent condition ($U = 998, p < 0.001; U = 277, p < 0.001$). This suggests that the congruency effect only holds if participants looked at the numbers while doing the number task, but not when they had to hold the two numbers in memory during the learning trial and then responded to the number task question. Congruency also did not impact the accuracy of the learning task ($U > 530, p > 0.9$) or the reaction time of the learning task ($U > 530, p > 0.7$). This indicates that the recruitment of inhibitory control did not impact reward learning.

Next, we checked the overall impact on accuracy in learning across conditions and experiments. We also compared differences in accuracy between conditions (Task-overlap vs. Control) across the 3 experiments. These differences capture the negative impact dual tasking had on learning performance. We found that the average difference in accuracy (capturing the effect of dual task) in experiment 1 was significantly greater than that in experiment 2 (0.180 vs. 0.126, $U = 326, p = 0.047$). This could be because the elongated inter-trial interval in experiment 1 made the

Single-task condition easier (Figure 2). Indeed, Wilcoxon test showed that the accuracy in the Single-Task condition was higher in experiment 1 than in experiment 2 ($U = 652.5, p = 0.016$).

To investigate whether task-switching, in the absence of dual-task, also impacted performance, we calculated the average difference in accuracy between the Task-switch and Task-overlap conditions in experiment 3. This difference was significantly different from 0 (0.074; $U = 44, p < 0.001$), indicating that the dual-task had a unique impact beyond task switching. However, the difference was significantly smaller than the average difference in accuracy in experiment 1 and in experiment 2 ($U = 343; 179, p = 0.047$; $p < .001$). Because the dual-task condition in experiment 1 and 2 were the same condition as the Task-overlap condition in experiment 3, this effect can only be explained by the fact that participants performed worse in the Task-switch condition in experiment 3, compared to the single-task condition in experiment 1 and 2. This illustrates a cost to task-switching vs. single task. Through a more direct comparison, we indeed found that participants performed worse in the Task-Switch condition in experiment 3 compared to the Single-Task conditions in experiment 1 and 2 ($U = 870.5, 752.5; p < 0.001, p = 0.002$).

Learning phase

We next sought to more carefully characterize condition and experiment effects using two-way ANOVA (see methods). Participants showed clear evidence of learning the stimulus-response mappings across all conditions. The probability of selecting the correct action increased with the number of stimulus appearances (Figure 2). Furthermore, learning was markedly weaker in the task-overlap condition than in the single-task condition in experiments 1 ($F(1, 180) = 86.27, p < 0.001$) and 2 ($F(1, 180) = 34.80, p < 0.001$) and the task-switch condition in experiment 3 ($F(1, 192) = 9.655, p = 0.002$) .

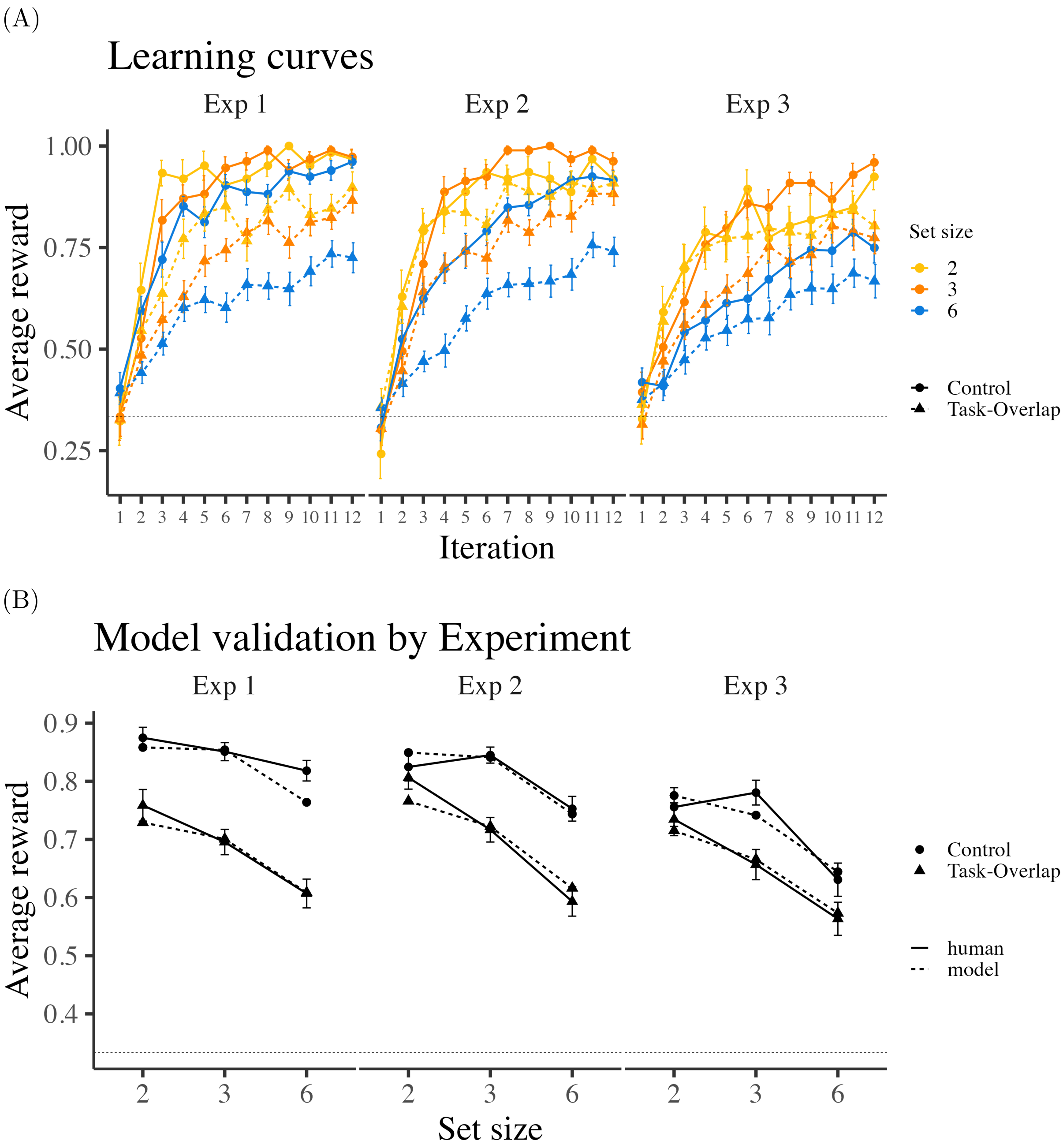

Figure 2. (A) Learning Curves: Participants learned stimulus-response associations over time, with significant effects of set size, experiment, and condition. Curves reflect the proportion of correct responses as a function of the number of times that each stimulus was presented, plotted separately for each set size and condition. (B) Model validation: the RLWM model captures well the overall proportion of correct choices across experiment, condition, and set size effects. Error bars reflect the standard error of the mean.

Regression analysis results confirmed that participants used both working memory and reinforcement learning processes to solve the task. Indeed, if working memory was recruited in this task, increasing the set size should decrease performance because holding more stimulus-response associations in mind across trials should make learning harder. We also analyzed the effect of cumulative reward for each stimulus in the regression model, obtained by adding all the points rewarded to each stimulus up to each trial. If reinforcement learning is incrementally increasing the value of the correct action associated with each stimulus, then performance should increase with the number of previous trials in which a stimulus has been rewarded. Replicating previous results (A. G. Collins et al., 2017; A. G. Collins & Frank, 2012), we observed both a significant negative effect of set size on performance in experiment 1 ($\beta = -0.175, \chi^2(5) = 25.433, p < 0.001$), experiment 2 ($\beta = -0.288, \chi^2(5) = 63.784, p < 0.001$), and experiment 3 ($\beta = -0.174, \chi^2(5) = 33.075, p < 0.001$), as well as a significant positive effect of cumulative reward on performance also in experiment 1 ($\beta = 0.874, \chi^2(5) = 478.120, p < 0.001$), experiment 2 ($\beta = 0.836, \chi^2(5) = 419.443, p < 0.001$), and experiment 3 ($\beta = 0.821, \chi^2(5) = 568.759, p < 0.001$), likely reflecting, respectively, the influences of working memory load and trial-by-trial reinforcement learning in this task (figure 2).

The regression model also tested the effect of "delay" on performance, captured by the number of trials passed since the last time a particular stimulus was observed and correctly responded to. We observed a significant negative effect of trial-based delay in experiment 1 ($\beta = -0.363, \chi^2(5) = 153.088, p < 0.001$), experiment 2 ($\beta = -0.399, \chi^2(5) = 167.838, p < 0.001$), and experiment 3 ($\beta = -0.366, \chi^2(5) = 161.038, p < 0.001$), suggesting that short-term forgetting occurs during the task (a result which is also consistent with the recruitment of working memory).

Finally, the regression results allowed us to consider dual task-overlap effects (figure 2). Consistent with our predictions, we observed a significant effect of condition in experiment 1 ($\beta = -0.605, \chi^2(5) = 313.310, p < 0.001$), experiment 2 ($\beta = -0.442, \chi^2(5) = 176.374, p < 0.001$), and experiment 3

($\beta = -0.188, \chi^2(5) = 49.701, p < 0.001$). Participants performed worse on the learning task in the Task-Overlap condition versus the Single-Task and Task-Switch condition. This result supports our prediction that performing the secondary task while concurrently retrieving and/or integrating reward feedback of the stimulus-response associations (Task-Overlap) had a stronger negative effect on learning relative to a situation where the secondary task is performed in between trials (Task-Switch) or relative to a situation where no secondary task was performed. Thus, actively maintaining information in WM affected instrumental learning performance.

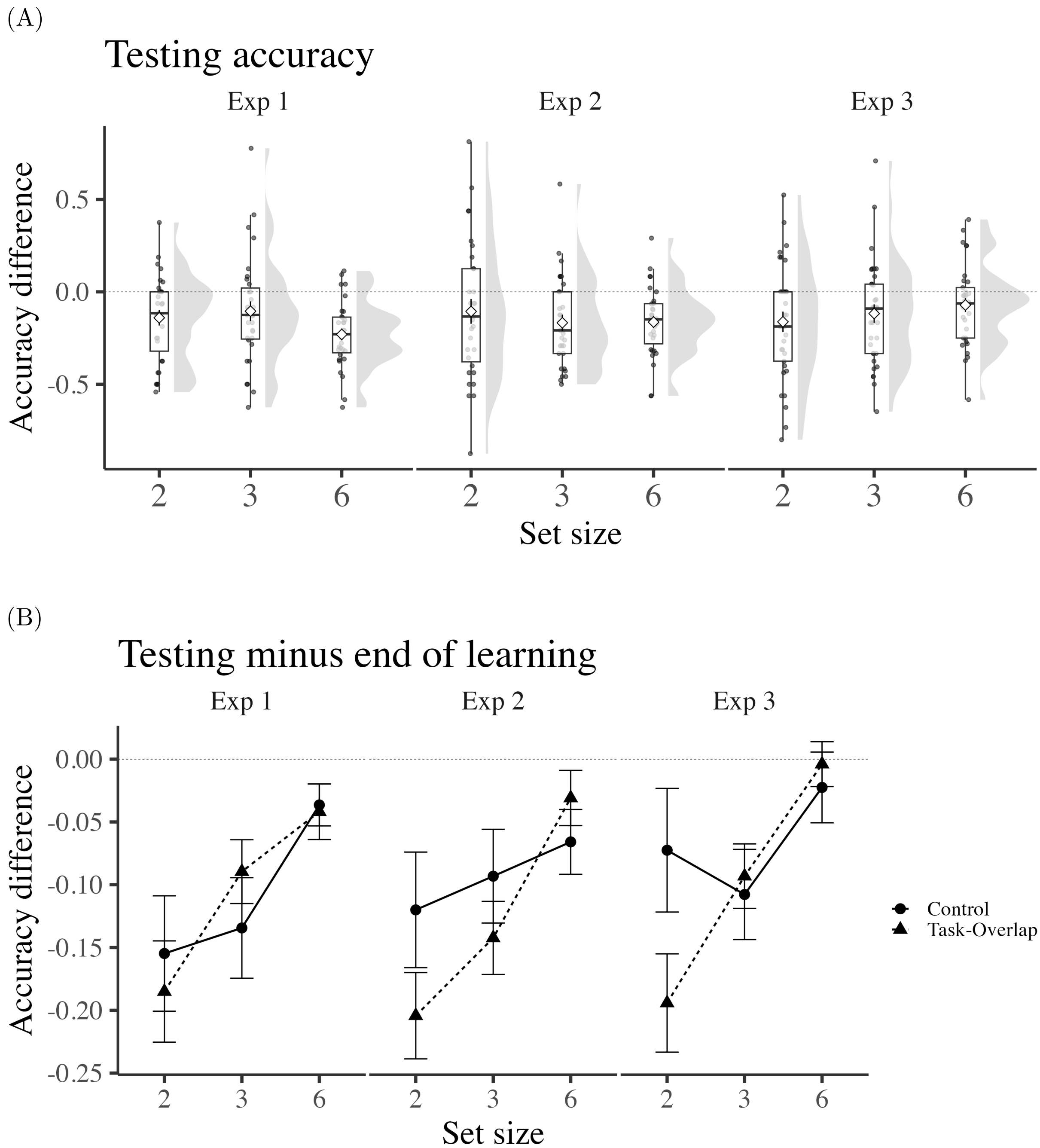

Figure 3. (A): The difference in testing phase accuracy: each dot represents the average accuracy of a participant in the Task-Overlap condition minus that in the the control condition. The diamond represents the mean of the average difference in accuracy. This shows that accuracy in testing phase was consistently lower in the task-overlap condition. (B): Change in accuracy: each point shows the mean value of the difference between the testing phase accuracy and the average accuracy of the last three (corresponding) learning trials. The difference in accuracy was not lower in the task-overlap condition, suggesting an impairment of the reinforcement learning system. Error bars reflect the standard error of the mean in both plots.

Testing phase

The asymptotic rate of learning performance, as expected, positively predicted the performance in the testing phase in experiment 1 ($\beta = 0.755, \chi^2(4) = 354.452, p < 0.001$), experiment 2 ($\beta = 0.781, \chi^2(4) = 380.872, p < 0.001$), and experiment 3 ($\beta = 0.908, \chi^2(4) = 528.983, p < 0.001$). This result gives more assurance that participants perform better on trials with stimuli that were well learned in the learning phase.

Next, we observed a significant *positive* effect of set size in experiment 1 ($\beta = 0.179, \chi^2(4) = 18.605, p < 0.001$), experiment 2 ($\beta = 0.192, \chi^2(4) = 21.247, p < 0.001$), and experiment 3 ($\beta = 0.161, \chi^2(4) = 17.154, p < 0.001$; figure 3). This finding replicates seemingly counter-intuitive previous findings (A. G. Collins, 2018): That is, this result suggests that when set size is low and working memory is contributing the lion's share to learning, long term retention of stimulus-action associations is actually hindered; conversely, when the set size is higher and reinforcement learning contributes more to learning, long-term retention is improved (even after controlling for asymptotic performance). Thus, the testing phase may act as a proxy for the strength of stimulus-response associations learned via the reinforcement learning system.

For the same reason, we might expect participants to potentially perform better in the testing phase on stimuli from the Dual-Task condition where working memory is directly taxed, assuming that the two systems (WM and RL) are competing. Contrary to this expectation, however, we found that participants performed worse in the testing phase on trials with stimuli from the Task-Overlap condition in experiment 1 ($\beta = -0.304, \chi^2(4) = 43.395, p < 0.001$) and experiment 2 ($\beta = -0.262, \chi^2(4) = 35.184, p < 0.001$). In experiment 3, participants performed worse on trials with stimuli from the Task-Overlap condition than the Task-Switch condition ($\beta = -0.142, \chi^2(4) = 13.466, p < 0.001$). This suggests that the effects of the condition we saw in the learning phase are not simply an effect on choices but actually on how well participants learned (figure 3). Otherwise, we would not see a condition-level effect on accuracy in the testing phase but only in the learning phase. This result also implies that directly blocking working memory seems to impair the performance of the reinforcement learning system as well, leading to decreased accuracy of testing phase responses.

Finally, we investigated the difference between the accuracy in the testing phase and the average accuracy of the last 3 trials of the learning phase (figure 3). We ran a linear mixed-effect regression on the difference in average accuracy with the following predictors: *condition*, *set size*, and *their interaction term*. While we replicated the previous finding (A. G. Collins, 2018) that the set size had a significant positive effect in experiment 1 ($\beta = 0.272, \chi^2(5) = 16.035, p < 0.001$), experiment 2 ($\beta = 0.242, \chi^2(5) = 14.285, p < 0.001$), and experiment 3 ($\beta = 0.249, \chi^2(5) = 12.960, p < 0.001$), we did not see a significant effect of condition ($\chi^2(5) < 1.874, p > 0.171$). This suggests that while the dual-task manipulation decreased participants' learning of the reward mapping, it did not affect the decay rate of the learning outcome.

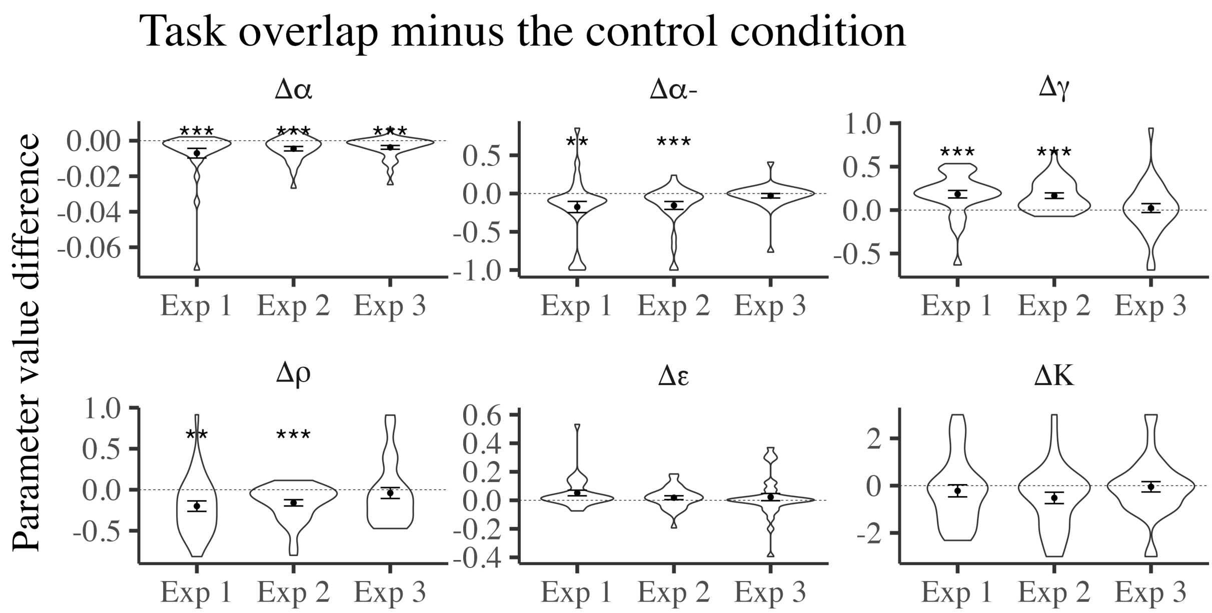

Figure 4. The difference in fitted parameter values between the Task-Overlap condition and the control condition: the learning rate of the reinforcement learning module (α) was consistently lower in the Task-Overlap condition, suggesting that dual-tasking impaired the reinforcement learning system. Outlier $|\Delta\alpha| > 0.2$ was removed for better visualization but included in the statistics reported. Error bars reflect the standard error of the mean. **:p<0.01, ***:p<0.001

Computational modeling results.

To directly investigate the mechanisms leading to condition and experiment effects, we next turned to RLWM modeling. We first looked at the model parameters we computed as a result of model fitting (figure 4). In both experiment 1 and experiment 2, we observed a significant difference between the Dual-Task and Single-Task conditions in the reinforcement learning rate α (Exp 1 $U = 71, p < 0.001$; Exp 2 $U = 71, p < 0.001$), learning bias α^- (Exp 1 $U = 101, p = 0.003$; Exp 2 $U = 80, p < 0.001$), forgetting γ (Exp 1 $U = 433, p < 0.001$; Exp 2 $U = 460, p < 0.001$) and basis working memory weight ρ (Exp 1 $U = 92, p = 0.002$; Exp 2 $U = 47, p < 0.001$). The fact that the dual-task manipulation did not have a significant effect on the ϵ noise parameter argues against the possibility that the effect of dual-tasking simply increased the noise of value-based choice without substantively impacting any executive function. Rather these results strongly suggest that dual-task manipulations during instrumental learning effectively interfere with both working memory itself, as suggested by decades of dual-task work, but also a putatively lower-level reinforcement learning system.

Interestingly, in Experiment 3, the only parameter value that significantly differed between the Task-Overlap and Task-Switch conditions was the reinforcement learning rate α ($U = 69, p < 0.001$). This result indicates that any dual task - whether it is one that is toggled between trials of learning or one that requires simultaneous memory maintenance during choice and updating – appears to hinder the reinforcement learning component of instrumental learning. On the other hand, the timing at which the extra memory load is imposed during dual-tasking appears to determine the severity of the dual-task effect on the reinforcement learning system. If the extra memory load is imposed during the value encoding stage of learning, as was the case in the Task-Overlap condition and all dual-task conditions in Experiments 1 and 2, we see a heightened hindrance of the reinforcement learning system.

Discussion

Many lines of evidence point to distinct processes contributing to instrumental learning (A. G. Collins & Frank, 2012; Daw, Gershman, Seymour, Dayan & Dolan, 2011; Lee, Seo & Jung, 2012). We have recently suggested that two of the processes, working memory (WM) and cortico-striatal reinforcement learning (RL), can be teased apart using specific task designs and computational modeling methods (A. G. Collins, 2018; A. G. Collins & Frank, 2012). One of the large gaps in this framework concerns the interaction of these two systems: whether one functionally depends on another. Our work provided one of the first direct evidence that the RL system indeed depends on executive functions because perturbing executive functions experimentally through the dual-task paradigm (Economides et al., 2015; Jiménez & Vázquez, 2005; Liefooghe, Barrouillet, Vandierendonck & Camos, 2008) led to worse learning outcome in the RL system, after controlling for direct contributions of WM to learning. We isolated the RL system using both an experimental method by introducing a test phase after the learning phase as well as through the "RLWM" model which was shown to nicely dissociate the separate contributions of WM and RL to instrumental learning (A. G. Collins, 2018).

The first main finding was that under the dual-task condition, participants performed significantly worse in the testing phase, where performance depended more on the information encoded in the RL system. Through modeling, we also found a clear effect of dual-tasking on the learning rate of the reinforcement learning system (Figure 4). That a tax on executive function would directly disrupt the primary parameter of the (putatively implicit, "lower-level") RL system is novel in our view, and may point to a deeper connection between executive function and RL than normally assumed (Rmus et al., 2021). We note that an alternative prediction could have been that the dual-task would disrupt the choice process itself, as opposed to learning-related processes. If that were the case, we would expect the noise (ϵ) parameter to be higher under dual-tasking, which we did not observe (Fig. 4). This further supports our interpretation that the dual-task interfered with learning computations, rather than choice per se.

Zooming out, we can interpret this result as an indication that working memory does not merely function as a separate storage system that works in parallel with the reinforcement learning system. If that were the case, we would expect taxing the executive function through dual-tasking to only disrupt the WM module while leaving the RL module unaffected. The fact that we found broader effects of the dual-task adds further support to the idea that there is a close dependency between WM and RL, as suggested in a recent similar study (A. G. Collins & Frank, 2018). We note that the dual-task paradigm does not allow us to directly speak to which specific component of executive function was responsible for the impairment of the RL system. We have a few speculations about why such impairment occurs. One hypothesis would be that greater noise in prefrontal representations, as expected from the addition of load, affects basic RL computations, for instance by disrupting "eligibility traces" that could be used to glue together states, actions, and rewards (Curtis & Lee, 2010) on short timescales. Another hypothesis is that some kind of explicit, internal verbal rehearsal process is being used by subjects in our task (Gershman, Markman & Otto, 2014), and that this process is disrupted or even blocked by the dual-task used in Experiments 1 and 2. Future work could use less verbalizable symbols in the dual-task to help tease out a role for verbal rehearsal here (Aspen H Yoo et al., 2023).

Our results also speak to some of the basic interpretations behind dual-tasking – dual-task manipulations are often thought to be useful tools for singularly taxing executive functions like attention and working memory, while sparing other (often sub-cortically linked) more implicit processes (Cohen, Ivry & Keele, 1990; Otto, Taylor & Markman, 2011; Vallesi, Arbula & Bernardis, 2014; Zeithamova & Maddox, 2006). While this general framework is useful and well-replicated, our results here complicate these assumptions somewhat, at least in the domain of instrumental learning. By showing that dual-tasking significantly disrupted a putatively non-cognitive RL system, we challenge the idea that dual-tasks leave implicit learning untouched (Rmus et al., 2021). Our findings may also have useful implications in more applied domains. For example in education, our findings suggests that factors that disturb executive functions (such as multi-tasking) may also impair more implicit learning mechanisms, like RL. In computational psychiatry, our findings highlight the difficulty of mapping mental disorders to specific sub-components of learning due to their mutual dependency. Overall, our findings point to a more general principle – seemingly distinct learning systems may often be at least somewhat intertwined, suggesting a more interactive approach to understanding learning (A. G. Collins & Frank, 2018; Fischer, Drosopoulos, Tsen & Born, 2006; McDougle, Ivry & Taylor, 2016).

Acknowledgments

We thank Helen Lu for help with data collection. We thank the CCN lab at UC Berkeley for helpful comments. This work was supported by NSF Grant 2020844 awarded to Anne G.E. Collins.

References

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. doi:10.1126/science.1736359

Cohen, A., Ivry, R. I. & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 17.

Collins, A., Ciullo, B., Frank, M. & Badre, D. (2017). Working memory load strengthens reward prediction errors. *Journal of Neuroscience*, *37*(16), 4332–4342. Epub 2017 Mar 20. doi:10.1523/JNEUROSCI.2700-16.2017

Collins, A. G. (2018). The Tortoise and the Hare: Interactions between Reinforcement Learning and Working Memory. *Journal of Cognitive Neuroscience*. doi:10.1162/jocn_a_01238

Collins, A. G., Ciullo, B., Frank, M. J. & Badre, D. (2017). Working Memory Load Strengthens Reward Prediction Errors. *The Journal of Neuroscience*, *37*(16), 4332–4342. doi:10.1523/JNEUROSCI.2700-16.2017

Collins, A. G. & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, *35*(7), 1024–1035. doi:10.1111/j.1460-9568.2011.07980.x

Collins, A. G. & Frank, M. J. (2018). Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. *Proceedings of the National Academy of Sciences*, *115*(10), 2502–2507. doi:10.1073/pnas.1720963115

Cortese, A., Lau, H. & Kawato, M. (2020). Unconscious reinforcement learning of hidden brain states supported by confidence. *Nature Communications*, *11*, 4429. doi:10.1038/s41467-020-17828-8

Curtis, C. E. & Lee, D. (2010). Beyond working memory: The role of persistent activity in decision making. *Trends in cognitive sciences*, *14*(5), 216–222.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S. & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature*, *378*(6554), 279–281. doi:10.1038/378279a0

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. doi:https://doi.org/10.1016/j.neuron.2011.02.027

Economides, M., Kurth-Nelson, Z., Lübbert, A., Guitart-Masip, M. & Dolan, R. J. (2015). Model-Based Reasoning in Humans Becomes Automatic with Training. *PLOS Computational Biology*, *11*(9), e1004463. doi:10.1371/journal.pcbi.1004463

Fischer, S., Drosopoulos, S., Tsen, J. & Born, J. (2006). Implicit learning–explicit knowing: A role for sleep in memory system interaction. *Journal of cognitive neuroscience*, *18*(3), 311–319.

Gabrieli, J. D. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology*, *49*, 87–115. doi:10.1146/annurev.psych.49.1.87

Gershman, S. J., Markman, A. B. & Otto, A. R. (2014). Retrospective revaluation in sequential decision making: A tale of two systems. *Journal of Experimental Psychology: General*, *143*(1), 182.

Jiménez, L. & Vázquez, G. A. (2005). Sequence learning under dual-task conditions: Alternatives to a resource-based account. *Psychological research*, *69*, 352–368.

Lee, D., Seo, H. & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual review of neuroscience*, *35*, 287–308.

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V. & Niv, Y. (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*, *93*(2), 451–463. doi:10.1016/j.neuron.2016.12.040

Liefooghe, B., Barrouillet, P., Vandierendonck, A. & Camos, V. (2008). Working memory costs of task switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 478.

Master, S. L., Eckstein, M. K., Gotlieb, N., Dahl, R., Wilbrecht, L. & Collins, A. G. (2020). Disentangling the systems contributing to changes in learning during adolescence. *Developmental cognitive neuroscience*, *41*, 100732.

McDougle, S. D., Ballard, I. C., Baribault, B., Bishop, S. J. & Collins, A. G. (2021). Executive function assigns value to novel goal-congruent outcomes. *Cerebral Cortex*, *32*(1), 231–247. doi:10.1093/cercor/bhab205

McDougle, S. D., Ivry, R. B. & Taylor, J. A. (2016). Taking aim at the cognitive side of learning in sensorimotor adaptation tasks. *Trends in cognitive sciences*, *20*(7), 535–544.

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A. & Wilson, R. C. (2015). Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms. *Journal of Neuroscience*, *35*(21), 8145–8157. doi:10.1523/JNEUROSCI.2978-14.2015

Otto, A. R., Taylor, E. G. & Markman, A. B. (2011). There are at least two kinds of probability matching: Evidence from a secondary task. *Cognition*, *118*(2), 274–279.

Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R. J. & Frith, C. D. (2008). Subliminal instrumental conditioning demonstrated in the human brain. *Neuron*, *59*(4), 561–567. doi:10.1016/j.neuron.2008.07.005

Posner, M. I. & Keele, S. W. (1967). Decay of Visual Information from a Single Letter. *Science*, *158*(3797), 137–139. doi:10.1126/science.158.3797.137

Rac-Lubashevsky, R., Cremer, A., Collins, A. G. E., Frank, M. J. & Schwabe, L. (2023). Neural index of reinforcement learning predicts improved stimulus-response retention under high working memory load. *Journal of Neuroscience*, *43*(17), 3131–3143. doi:10.1523/JNEUROSCI.1274-22.2023

Rescorla, R. (1972). A theory of Pavlovian conditioning : Variations in the effectiveness of reinforcement and nonreinforcement.

Rmus, M., He, M., Baribault, B., Walsh, E. G., Festa, E. K., Collins, A. G. & Nassar, M. R. (2023). Age-related differences in prefrontal glutamate are associated with increased working memory decay that gives the appearance of learning deficits. *eLife*, *12*, e85243. doi:10.7554/eLife.85243

Rmus, M., McDougle, S. D. & Collins, A. G. (2021). The role of executive function in shaping reinforcement learning. *Current Opinion in Behavioral Sciences*, *38*, 66–73. doi:10.1016/j.cobeha.2020.10.003

Shohamy, D. (2011). Learning and motivation in the human striatum. *Current Opinion in Neurobiology*, *21*(3), 408–414. doi:https://doi.org/10.1016/j.conb.2011.05.009

Sinclair, A. H., Wang, Y. C. & Adcock, R. A. (2023). Instructed motivational states bias reinforcement learning and memory formation. *Proceedings of the National Academy of Sciences*, *120*(31), e2304881120.

Sutton, R. S. & Barto, A. G. (1998). *Introduction to reinforcement learning*. MIT press Cambridge.

Vallesi, A., Arbula, S. & Bernardis, P. (2014). Functional dissociations in temporal preparation: Evidence from dual-task performance. *Cognition*, *130*(2), 141–151.

Viejo, G., Khamassi, M., Brovelli, A. & Girard, B. (2015). Modeling choice and reaction time during arbitrary visuomotor learning through the coordination of adaptive working memory and reinforcement learning. *Frontiers in Behavioral Neuroscience*, *9*. doi:10.3389/fnbeh.2015.00225

Wilkinson, L. & Jahanshahi, M. (2007). The striatum and probabilistic implicit sequence learning. *Brain Research*, *1137*, 117–130. doi:https://doi.org/10.1016/j.brainres.2006.12.051

Wilson, R. C. & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, e49547. doi:10.7554/eLife.49547

Yoo, A. & Collins, A. (2022). How working memory and reinforcement learning are intertwined: A cognitive, neural, and computational perspective. *Journal of Cognitive Neuroscience*, *34*(4), 551–568. doi:10.1162/jocn_a_01808

Yoo, A. H. [Aspen H], Keglovits, H. & Collins, A. G. (2023). Lowered inter-stimulus discriminability hurts incremental contributions to learning. *Cognitive, Affective, & Behavioral Neuroscience*, *23*(5), 1346–1364.

Zeithamova, D. & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & cognition*, *34*(2), 387–398.