Reward learning and working memory: effects of massed versus spaced training and post-learning delay period

Abbreviated title: Massed versus spaced learning

G. Elliott Wimmer[1,2] and Russell A. Poldrack[2]

[1] Max Planck University College London Centre for Computational Psychiatry and Ageing Research and the Wellcome Centre for Human Neuroimaging, University College London, London, UK

[2] Department of Psychology, Stanford University, 450 Jane Stanford Way, Stanford, CA, 94305

Corresponding author:
G. Elliott Wimmer
Max Planck UCL Centre for Computational
Psychiatry and Ageing
Wellcome Trust Centre for NeuroImaging
10-12 Russell Square
University College London
e.wimmer@ucl.ac.uk
+44 07475646324

Acknowledgments

The authors thank Jamie Li for assistance in data collection and helpful discussions. Research was supported by NIH R01AG041653 [RP] and a research fellowship from the Deutsche Forschungsgemeinschaft [GEW].

Abstract

Over the past few decades, neuroscience research has illuminated the neural mechanisms supporting learning from reward feedback, demonstrating a critical role for the striatum and ascending midbrain dopamine system. However, in humans, short-term working memory can also play an important role in reward-based learning, particularly in common paradigms that are relatively compressed in time. Given the growing use of reward-based learning tasks in translational studies in computational psychiatry, it is important to further investigate the degree to which apparent reinforcement learning is driven by working memory contributions. In our experiments, we manipulated the spacing between repetitions and also the post-learning delay period duration. Supporting a significant role for working memory in reward-based learning, we found that learning was faster for massed-trained compared to spaced-trained stimuli, and that brief interruptions of massed learning led to drops in subsequent performance. Further, individual differences in working memory capacity positively correlated with massed learning performance. Critically, after a delay period but not immediately, relative performance decayed in the massed condition and increased in the spaced condition. Overall, our results provide further support that working memory plays a large role in temporally condensed reward learning designs. We suggest that spacing training across or between sessions is a promising direction to pursue in future studies investigating the mechanisms of reward-based learning, and particularly in translational research.

Introduction

The accumulation of rewarding and aversive experiences exerts a strong influence on later decision making. When making a choice between an apple and a banana, for example, our decision relies on values shaped by many experiences spread across years. In general, repeated experiences of stimulus- and action-reward associations in our lives are often separated by minutes, hours, days, or even longer, and recent research has shown that spaced training leads to value associations that are resistant to forgetting, similar to habits (Kim et al., 2015; Wimmer et al., 2018; van de Vijver and Ligneul, 2019). However, there is a striking difference between the proposed slow, gradual learning mechanism that supports the learning of habitual stimulus- and action-value associations (Yin and Knowlton, 2006) and actual experimental designs that are commonly used to study this learning mechanism in humans. Such designs feature closely-spaced repetitions and often include fluctuating values (e.g. Daw et al., 2011; Wimmer et al., 2014).

Critically, recent research has begun to illuminate how performance in such "massed" training contexts can be supported by short-term memory processes in addition to gradual stimulus- and action-reward learning processes supported by dopaminergic effects in cortico-striatal circuits (Collins and Frank, 2012; Wimmer et al., 2018). Reward learning paradigms are increasingly being extended to study potential learning dysfunctions in mood and psychiatric disorders as well as addiction (Maia and Frank, 2011; Montague et al., 2012; Huys et al., 2016; Moutoussis et al., 2016). Thus, it is important to determine whether common lab-based reward learning studies successfully isolate the learning processes of interest. For example, it has been reliably demonstrated that apparent deficits in reward-based learning in patients with

schizophrenia are better accounted for by a deficit in working memory (Collins et al., 2014; Collins et al., 2017).

While understanding interactions between massed training and working memory has been the focus of a number of recent studies, the inverse of massed training – spaced training – has been relatively under-explored (Wimmer et al., 2018). Spacing between learning events is well-known to have benefits for verbal memory, motor skill learning, and education (reported in Ebbinghaus, 1913; Lee and Genovese, 1988; Donovan and Radosevich, 1999; Janiszewski et al., 2003; Cepeda et al., 2006). In animal research on reward-based learning, increased spacing between repetitions reliably increases learning speed (Teichner, 1952; Terrace et al., 1975).

Humans are likely to use short-term working memory to support massed learning, so we would not expect to find similar increases in ongoing performance with greater spacing. In verbal and motor training, in fact, it has been found that spaced training often leads to lower performance during training but better retention of learning at a later test (Schmidt and Bjork, 1992; Taylor and Rohrer, 2010; Soderstrom and Bjork, 2015). A benefit of spaced training reflected in later measures of learning may arise from multiple mechanisms, including offline replay that strengthens reward associations (Sutton, 1990; Olafsdottir et al., 2018). It is unknown whether processes operating during post-learning rest may benefit relatively simple stimulus-reward association learning in humans.

Our experiments examine the relationship between short-term memory and learning in conditions when learning repetitions are massed versus spaced. We developed a single experimental paradigm that could provide multiple measures of this relationship and simultaneously examine the effect of spacing on performance during

and after learning. During learning, abstract stimuli were probabilistically paired with rewards or losses that depended on the participant's response. Training for a given "massed-trained" stimulus was completed in less than a minute, while training for a given "spaced-trained" stimulus was spread across approximately 15 min. All repetitions of a massed stimulus were presented on sequential trials, with occasional spaced stimulus trials inserted in-between these repetitions. Both massed and spaced stimuli were experienced an equal number of times. Following learning, participants engaged in a choice test phase either immediately (the No Delay condition) or after approximately 15 minutes (the Delay condition).

We predicted that 1) Massed training would lead to better performance during training; 2) Brief interruptions to massed training would be followed by performance decreases; 3) Working memory capacity would positively relate to massed but not spaced learning performance; and 4) Learning for massed- but not spaced-trained stimuli would decrease following a rest period. Further, the post-learning rest delay also allowed us to ask whether a delay would lead to relative improvements in memory for spaced-trained value associations.

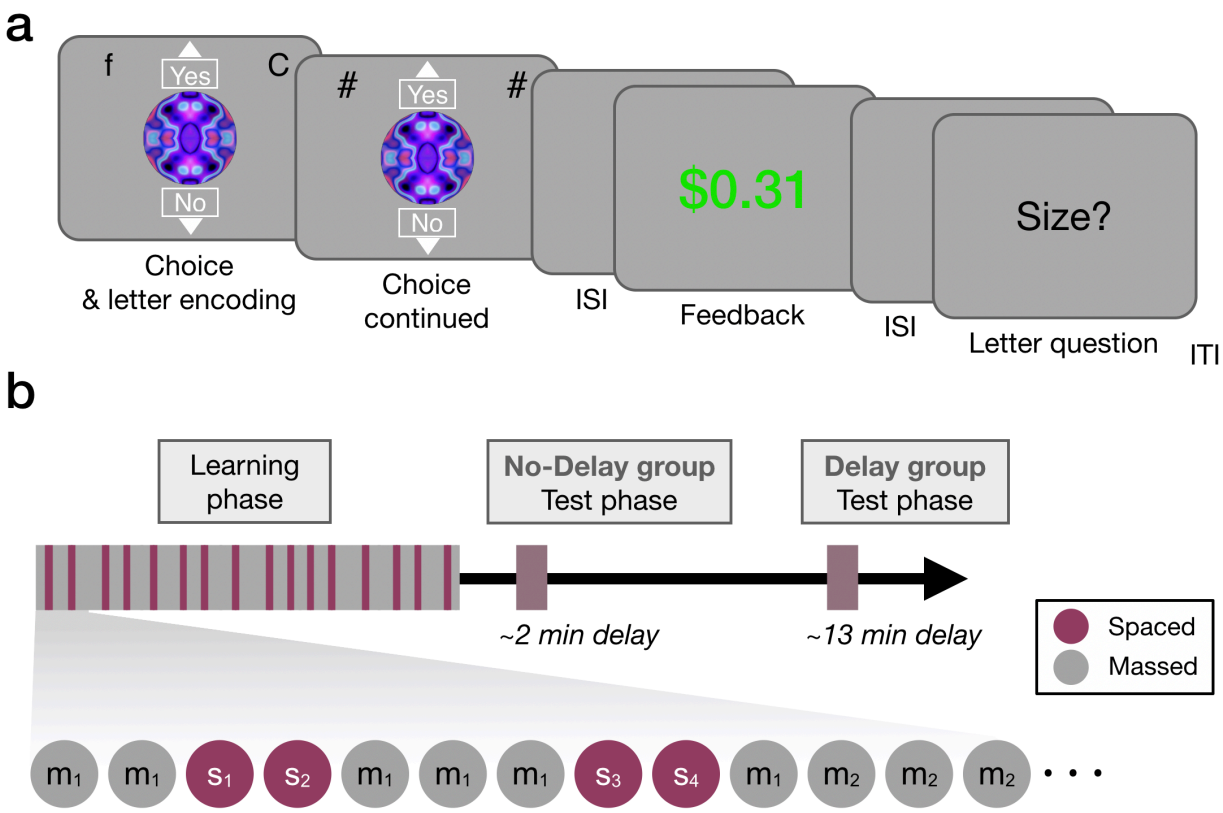

Figure 1. (**a**) Reward learning phase. Participants encoded 2 letters for the secondary task and then made a 'Yes' or 'No' response to the abstract circle stimulus. Following reward feedback, participants responded to a question about the letters. (**b**) Experimental design schematic. Timeline of learning and test phases for the No-Delay group and the Delay group. Magnification: an example trial-by-trial progression of massed- and spaced-trained stimuli. Repetitions of massed-trained stimuli proceeded sequentially (e.g. m_1 followed by m_1; grey circles), with occasional interruptions by spaced-trained stimulus trials (e.g. s_1, s_2, s_3, s_4; magenta circles).

Methods

Participants

Participants were recruited via advertising on the Stanford Psychology paid participant pool web portal. Informed consent was obtained in a manner approved by the Stanford

University Institutional Review Board. Target sample size was based on a related study of spacing and reward-based learning (Wimmer et al., 2018). In the Delay group (Experiment 1), from an initial number of 35 participants, 8 were excluded based on performance according to the methodology described below, leaving a total of 27 participants (13 female; mean age = 22.2, range: 19-29). In the No-Delay group (Experiment 2), from an initial number of 34 participants, 7 were excluded based on performance, leaving a total of 27 participants (15 female; mean age = 21.7; range: 18-30). Participants were paid $10/hour for the approximately 75-minute experiment, plus monetary rewards from the learning phase and choice test phase, leading to total compensation of approximately $20.

Participants were excluded based on three separate learning phase performance measures unrelated to our hypotheses of interest: poor learning performance for massed stimuli, extreme response bias in spaced learning, and poor dual-task letter performance. We did not use the rating or choice test phases for exclusion. First, we excluded participants with mean massed learning performance on the last two repetitions (the 6th and 7th) of less than 60% (1 exclusion in the Delay group; 5 in the No-Delay group). Near-chance performance in the massed condition is likely related to poor overall attention to the reward learning task. Second, we excluded participants with an extreme spaced-stimulus Yes/No response bias in the second half of repetitions (5-7), where $\geq$ 90 % of responses were of the same option across all spaced stimuli (3 exclusions in the Delay group; 1 in the No-Delay group). Strong response bias indicates an overreliance on a fixed strategy for the spaced stimulus trials, which may interfere with non-strategic stimulus-specific learning mechanisms. Third, we excluded participants with poor performance on the letter judgment secondary task (4 exclusions

in the Delay group; 2 in the No-Delay group). Near-chance performance indicates a lack of attention to the secondary task. As performance in the dual-task was lower than expected overall, we used an exclusion threshold $\leq$ 60% accuracy. Raising the performance exclusion threshold for the letter task did not qualitatively affect the results.

Experimental Design

The reward learning task included three phases: a learning phase, a rating test phase, and a choice test phase. The No-Delay and Delay groups differed only in a potential delay between the end of the learning phase and the beginning of the test phases. In the Delay group (Experiment 1), an approximately 13 min break followed before the first rating phase trial. In the No-delay group (Experiment 2), an approximately 2 min delay followed the last learning trial and the first rating phase trial (due to the delay in the experimenter re-entering the testing room and providing instructions for the test phase). The rest period in the Delay group was intended 1) to provide a pause during which potential forgetting as well as offline replay processes might affect value associations (Gershman et al., 2014), and 2) to better equalize the relative delays between learning and test for the massed stimuli. During this time, participants were told that they were to take a break and could occupy themselves by leaving the testing room, interacting with their mobile devices, or browsing the internet on the testing room computer.

The learning phase was similar to a simple reward-based learning task used previously (Gerraty et al., 2014; Wimmer et al., 2018). Abstract circle stimuli were associated with either potential reward or potential loss, depending on the participant's response. The goal was to learn the best response for each stimulus (arbitrarily labeled

“Yes” and “No”), in order to win as much as possible in this phase. Participants used up and down arrow keys to make “Yes” and “No” responses, respectively. The optimal response for reward-associated stimuli was to select “Yes” in order to win the potential gain (mean +$0.25, in green font; 86 % probability of success) vs. the alternative small loss (mean -$0.05, in grey) for a “No” response. The optimal response for loss-associated stimuli was to select “No” in order to achieve a non-loss outcome (mean +$0.05, in grey; 86 % probability of success) and avoid the loss (mean +$0.05, in red) associated with a “Yes” response. Repetitions of individual massed stimuli were presented in sequential trials, with occasional spaced stimulus trials intermixed, as described below.

In a single reward learning trial, a stimulus was first presented with the options “Yes” and “No” above and below the image, respectively (**Fig. 1B**). Above the circle, two letters were presented on the left and right sides of the screen for the concurrent short-term memory task. One letter was upper-case (large), while the other was lower-case (small). Additionally, one letter was from earlier in the Latin alphabet while one was from later in the alphabet. The letters appeared on the screen for 0.30 s before being replaced by pound signs (“#”) for 0.20 s. The participants could make their Yes/No response to the circle stimulus at any time during the letter presentation phase or afterwards using the up and down arrow keys on the keyboard. Participants had 2 s overall to make their response to the circle. After a response, the Yes/No options remained on the screen for the remainder of the full choice period. A 1 s blank inter-stimulus interval followed. Next, reward feedback was shown in text in the center of the screen for 1.5 s. If a response was not made in the choice period, participants were shown text stating “Too late or wrong key! - $0.50” in red font. A brief 0.25 s blank inter-

stimulus interval followed. Next, either “Earlier?” or “Size?” text was shown on the screen, referring to the letters shown initially during the choice (which were no longer on the screen). Here, participants decided which letter (left or right) was earlier in the alphabet or capitalized, respectively. Participants had 2 s to make a response, using the left and right arrow keys. A blank ISI of 0.50 s followed the response. If an incorrect response was made or if no response was recorded, “Incorrect!” appeared on the screen in red font for 0.75 s. If a correct response was made, a brief fixation of 0.25 s followed. Finally, an inter-trial-interval including a white fixation cross was presented for an average of 2 s (range: 0.50-3.25 s), followed by a trial-start indicator where the white fixation changed to black for 0.25 s. The learning phase began with 12 practice trials, including one reward- and one loss-associated practice stimulus, during which the letter presentation period was increased by 0.50 s and the choice response period was increased by 2 s. Mid-way through the practice block and approximately every 25 trials thereafter, the computer displayed a warning if letter task performance fell below 66%.

Participants accumulated the most rewards by learning to make the correct “Yes” response for reward-associated stimuli and “No” response for loss-associated stimuli. The correct response led to the best outcome ~86% of the time. Displayed reward amounts were jittered by adding a uniform distribution of ±5 cents around the means. The massed condition included 4 reward- and 4 loss-associated stimuli; the spaced condition included 3 reward- and 3 loss-associated stimuli. Each stimulus was repeated 7 times, yielding a total of 98 trials, excluding practice. All repetitions of an individual massed stimulus were in sequence, with occasional interruptions by 1-4 spaced stimulus trials (mean 2.73), with all repetitions for a given massed stimulus spread across ~1 min. The 7 repetitions of each spaced stimulus were spread across the full

learning phase of ~15 min. For the spaced stimuli, the 6 stimuli were presented a single time in a pseudo-random order per each of the 7 repetitions, with no direct repetitions of the same stimulus. The average total trial separation (including massed and spaced stimuli) between repetitions of spaced-trained stimuli was 14.2 trials, with a minimum of 5 and a maximum of 28. The order of the presentation of each of the massed stimuli was pseudorandomized to avoid long stretches of a reward or loss context, such that no more than two massed stimuli associated with reward (or loss) followed in sequence. The order of the spaced stimuli was also pseudorandomized, with a set of spaced trials beginning with an approximately equal number of reward and loss stimuli amidst a reward or loss massed stimulus sequence. 5 of the 7 transitions between different massed stimuli were marked by the presence of one or more spaced trials. Stimulus assignment to the massed and spaced condition was counterbalanced across participants.

The letter task stimuli were taken from the set of letters ‘a’ through ‘j’, excluding ‘i’. Letters per trial were balanced such that the larger or earlier letter appeared approximately equally on the left and right side of the screen, leading to an approximately even distribution of the correct response to the left and right options. On ~85% of trials, the correct answers for the two potential probe questions were different; on the remaining trials, one letter was both earlier and a capital letter. The stimuli and queries were pseudorandomized so that queries and correct responses were approximately balanced across spaced and massed trials.

After the learning phase, participants completed a reward rating phase and an incentive-compatible choice phase. As noted above, in the Delay group, participants first had a break of ~13 min after the learning phase (mean 12.76 min, range 7.6-22.5

min). In the No-Delay group, the time between the last learning trial and the first rating trial was ~2 min (mean 2.04 min, range 1.4-5.0 min), allowing for the experimenter to return to the testing room and administer instructions. In the reward rating phase, participants saw each abstract circle stimulus and tried to recall whether that stimulus was associated with reward or loss. Below the stimulus, a rating scale appeared, anchored by “0% reward” on the left and “100% reward” on the right. Participants were instructed to try to remember the value of a stimulus, using their best guess or gut feeling. They were instructed that the endpoints of the scale represented complete confidence in their answer, while points closer to the middle indicated lower confidence. Participants indicated their response using a computer mouse (with no time limit), followed by a 3 s ITI. The phase began with a single practice trial followed by the massed and spaced stimuli in a pseudo-random order.

Next, participants completed in the incentive-compatible choice test phase. On each trial, one stimulus was presented on the left and an alternative stimulus was presented on the right, with location randomized. Participants were instructed to choose the stimulus that they thought had been associated with reward over the stimulus they thought had been associated with loss. Further, participants were informed that they would not receive feedback, but that choice of the reward-associated stimulus would lead to a higher additional monetary reward at the end of the experiment. As in the rating phase, participants were instructed to use their best guess or gut feeling. Participants made their responses using a 4-point scale: “(1) Sure left, (2) Guess left, (3) Guess right, (4) Sure right” using the 1-4 keys. They were further instructed that the level of confidence of their answer did not affect the potential bonus for correct choices. A 3.5 s ITI followed the response. The phase began with a single practice trial using the

two stimuli from the practice learning trials followed by the choice trials in pseudo-random order.

The primary learning measure in the choice test phase was based on choices that contrasted a reward vs. a loss-associated stimulus from within the spaced condition or within the massed condition. The phase also contained a secondary kind of choice comparing a stimulus from the spaced to the massed condition where both had the same value association (both reward or both loss). In the Delay group, 43 choices were presented. All of the potential combinations of reward versus loss stimuli from the spaced condition were presented (9 choices). A subset of the massed condition choices were presented (10 choices); a subset was used in order to reduce potential fatigue. Choice order was pseudo-randomized. The first 8 choices were critical choices (within spacing condition) while the last 12 choices were between-spacing condition choices. The choice test phase in the No-Delay group used the same set of choices as the Delay group, plus the additional 6 massed-massed choices that were held out in the No-Delay group, yielding 49 choice trials. After the choice test phase, participants completed a short written questionnaire.

In the No-Delay group (Experiment 2), we additionally administered a working memory measure, the operations span task (O-SPAN) (Lewandowsky et al., 2010; Otto et al., 2013). In the O-SPAN, participants made accuracy judgments about simple arithmetic equations (e.g. ‘2 + 2 = 5’). After a response, an unrelated letter appeared (e.g. ‘B’), followed by the next equation. Arithmetic-letter sequences ranged in length from 4 to 8. At the completion of a sequence, participants were asked to type in the letters that they had seen in the original order, with no time limit. Each of the sequence lengths was repeated 3 times with different equations and letters in a pseudo-random

order. In order to ensure that participants were fully practiced in the task before it began, the task was described in-depth in instruction slides, followed by 5 practice trials. Scores were calculated by summing the number of letters in fully correct letter responses across all 15 trials (mean, 50.7; range, 19-83) (Otto et al., 2013; Wimmer et al., 2018). All participants maintained a level of correct arithmetic performance above 70%, with group mean performance of 94%.

Behavioral analysis

Behavioral analyses were primarily conducted in Matlab 2016a (The MathWorks, Inc., Natick, MA). Learning performance was quantified as percent correct choice ("Yes" for the reward-associated stimuli, and "No" for the loss-associated stimuli) and compared to chance using a t-test. To examine the effect of interspersed spaced trials on concurrent massed learning performance, the performance change from pre- to post-interruption was compared to a balanced control performance change across repetitions with no interruption. This control measure was constructed by computing an average of performance changes weighted by the actual number of times that massed learning was interrupted at a given learning repetition. In this way, the balanced control measure included changes from repetition 3 to 4 and repetition 4 to 5 with a smaller weight given to changes from repetitions 2 to 3, 5 to 6, and 6 to 7.

Test phase choices were averaged within condition. We also compared the end of learning performance with choice performance between the massed- and spaced-trained conditions, using the last repetition from the learning phase. Note that while learning phase accuracy and choice accuracy use different behavioral measures, any interaction between change in performance and delay condition only depends on

relative differences in performance, independent of the different underlying measures. Test phase reward ratings were recorded in a graded scale from 0 to 100, where 50 indicated neutral. Regression analyses examining the relationship between reward ratings (with one value per stimulus) and test phase choices were conducted within-participant. At the second level, coefficients across the group were compared versus zero using a t-test.

Learning phase multilevel regression analyses were conducted in R (https://www.r-project.org/) using the lme4 package. All models entered subject as well as intercepts for variables of interest as random effects. The implementation of the multilevel models followed previous procedures (Braun et al., 2018). We used the glmer function to assess the relationship between variables of interest and performance or reaction time. To ensure convergence, models were run using the bobyqa optimizer set to 10^6 iterations. We estimated confidence intervals using the confint.merMod function and p-values using the bootMer function (both from the lmer4 package) using 2500 iterations. All reported p-values are two-tailed.

We additionally examined correlations between working memory capacity (as measured with the O-SPAN) and behavioral performance. Individual differences in performance from the learning phase were based on performance after sufficient task exposure (here, the second half of the learning phase) in order to prevent confounding factors such as initial task adjustment, attentional orienting, and task-set learning from adding noise to a potential relationship with working memory, following previous procedures (Wimmer and Poldrack, 2018). Correlations were computed using Pearson's correlation. Statistical comparison of the difference between correlations was computed using Steiger's test for differences in dependent correlations.

For all results of interest, we tested whether non-significant results were weaker than a moderate effect size using the Two One-Sided Test (TOST) procedure (Schuirmann, 1987; Lakens, 2017) as implemented in the TOSTER library in R (Lakens, 2017). We used bounds of Cohen's *d* = 0.57, where power to detect an effect in the included group of n = 27 participants in either the No-Delay or Delay group is estimated to be 80%. For correlations, the *r*-value cutoff for 80% power was estimated to be 0.37.

Data availability

The full behavioral dataset is available at OSF (https://osf.io/xu48j/).

Results

Learning. In the learning phase, participants learned the best response for sequentially presented abstract circle stimuli. For reward-associated stimuli, participants could gain a reward most of the time by selecting the “Yes” option; for loss-associated stimuli, participants could avoid a loss most of the time by selecting the “No” option. All repetitions of a given massed-trained stimulus were completed within 1 minute, while repetitions of “spaced” stimuli were interleaved between massed trials, such that all repetitions of a given spaced-trained stimulus were spread across approx. 15 minutes. Thus, spacing was manipulated within-participant and within the same session, in contrast to a multi-session spacing paradigm we reported recently (Wimmer and Poldrack, 2018).

We examined massed versus spaced learning in two groups of participants to examine the maintenance of learning across a post-learning delay. In the No-Delay group the test phases proceeded after the end of learning, while in the Delay group, the test phases proceeded after a break of approximately 13 min. We predicted that a post-learning delay may have different effects on later performance for massed- versus spaced-trained stimuli.

In the No-Delay group, performance increased across training, as demonstrated by an effect of stimulus repetition on performance (multilevel model regression β = 0.154 [0.113 0.198]; z = 7.279, p < 0.001; **Fig. 2a**). On average, for massed-trained stimuli, performance increased from 46.3 % (CI [38.9 53.6]) to 82.0 % (CI [76.9 87.1]) by the last (7th) repetition. For the spaced condition, performance increased from 48.4 % (CI [41.4 55.4]) to 57.2 % (CI [50.1 64.2]) by the last repetition.

We expected that performance for the immediately-repeating massed-trained stimuli would be assisted by short-term memory processes such as working memory. Supporting this prediction, we found that performance for massed-trained stimuli was significantly higher than performance for spaced-trained stimuli at the end of training ($t_{(26)}$ = 6.91, CI [11.3 20.8]; $p < 0.001$; t-test). This massed benefit was also reflected in a stronger effect of repetition in the massed versus spaced condition (β = 0.197 [0.090 0.312]; z = 3.435, $p < 0.001$). The effect of repetition within the massed condition alone was significant, while the effect in the spaced condition was positive but non-significant (massed condition β = 0.257 [0.189 0.330]; z = 7.167, $p < 0.001$; spaced condition β = 0.061 [0.001 0.123]; z = 1.925, p = 0.058).

The learning phase procedure was the same across the No-Delay and Delay groups. As expected, learning performance was similar across the groups. Performance in the Delay group increased across training (effect of stimulus repetition, β = 0.197 [0.154 0.241]; z = 9.093, $p < 0.001$; **Fig. 2a**). On average, in the massed condition performance increased from 47.9 % (CI [42.8 53.1]) to 89.2 % (CI [84.7 93.7]) by the last repetition. In the spaced condition, performance increased from 45.5 % (CI [38.4 52.6]) to 57.9 % (CI [50.1 65.7]) by the last repetition. Again reflecting a benefit of working memory on massed learning, performance was significantly higher in the massed versus the spaced condition ($t_{(26)}$ = 7.14, CI [12.7 22.9]; $p < 0.001$). Further, the effect of training repetition was significantly stronger in the massed condition (β = 0.225 [0.126 0.331]; z = 4.470, $p < 0.001$). Nevertheless, the effect of repetition in the spaced condition alone was also significant (β = 0.10 [0.039 0.164]; z = 3.240, p = 0.0008; massed condition β = 0.320 [0.250 0.391]; z = 8.790, $p < 0.001$).

As the learning phase progressed, learning performance for massed-trained

stimuli presented later numerically increased (No-Delay group: β = 0.046 [-0.003 0.096]; z = 1.795, p = 0.066; Delay group: β = 0.085 [0.021 0.152]; z = 2.573, p = 0.016), an effect consistent with a positive relationship between task exposure and learning of task set on performance. Separately, in both the No-Delay and Delay groups, we also found that performance was higher for reward- versus loss-associated stimuli (No-Delay: β = 0.621 [0.198 1.042]; z = 2.939, p = 0.0033; Delay: β = 0.651 [0.351 0.947]; z = 4.411, p < 0.001).

During the learning phase, to keep massed performance below ceiling, participants engaged in a demanding secondary task which required them to remember letter identity and letter size for two letters over the course of each trial (e.g. 'B' and 'e'; **Fig. 1a**). In the No-Delay group, performance on the letter task was far above chance (massed, 81.5 % CI [78.2 84.8]; spaced, 80.1 % CI [76.4 83.7]) and did not differ between the spacing conditions ($t_{(26)}$ = 1.024, CI [-1.4 4.3]; p = 0.315; TOST p = 0.032). Similarly, in the Delay group, letter performance was far above chance (massed, 74.9 % CI [71.3 78.6]; spaced, 76.0 % CI [71.8 80.3]) and did not differ between the spacing conditions ($t_{(26)}$ = -0.699, CI [-1.4 4.5]; p = 0.491; TOST p = 0.016).

Independent of spacing, we found that performance on the letter task in both groups was higher for reward-associated stimuli than loss-associated stimuli (No-Delay reward 84.7 % CI [81.6 87.8]; loss 77.1 % CI [72.9 81.2]; $t_{(26)}$ = 4.33, CI [4.0 11.3]; p < 0.001; Delay reward 78.3 % CI [74.0 82.6]; loss 72.5 % CI [68.6 76.3]; $t_{(26)}$ = 3.04, CI [1.9 9.8]; p = 0.006). The difference between letter task performance on reward- versus loss-associated stimulus trials tended to be greater in the massed- than spaced-trained condition (No-Delay $t_{(26)}$ = 2.00, CI [-0.2 12.8]; p = 0.056; Delay $t_{(26)}$ = 6.61, CI [10.3

19.6]; $p < 0.001$). The performance benefit on the secondary letter task for reward-associated stimuli trials mirrors the learning performance benefit for reward-associated stimuli, suggesting that higher learning performance may release cognitive resources that could be allocated to the secondary letter task.

Effect of interruptions on massed learning performance

Our second test of the prediction that performance on massed-trained stimuli would be assisted by short-term memory examined the effect of interruptions of massed stimulus repetitions by occasional interleaved spaced stimulus trials (**Fig. 1b**). Across the repetitions of a given massed stimulus, spaced stimulus trials were pseudo-randomly inserted (median 2 spaced trials per interruption, range 1-3), and we expected that interruptions would lead to a transient decrease in massed stimulus performance.

Supporting this prediction, in the No-Delay group, we found that the interruption of massed stimulus repetitions by spaced trials negatively affected post-interruption performance (post-pre interruption performance versus balanced no-interruption change over consecutive repetitions; $t_{(26)} = 2.318$, CI [0.009 0.141]; $p = 0.029$; **Fig. 2b**). The effect of interruption was also reflected in reaction time, such that reaction time exhibited a transient increase post-interruption (post-pre interruption performance versus balanced no-interruption change over consecutive repetitions; $t_{(26)} = -3.015$, CI [-0.121 -0.023]; $p = 0.006$). In the Delay group, we also found a negative effect of interruption of massed learning ($t_{(26)} = 5.280$, CI [0.123 0.280]; $p < 0.001$; **Fig. 2b**). The effect of interruption was again reflected in reaction time, with reaction time relatively increased post-interruption ($t_{(26)} = -3.571$, CI [-0.146 0.039]; $p = 0.0014$).

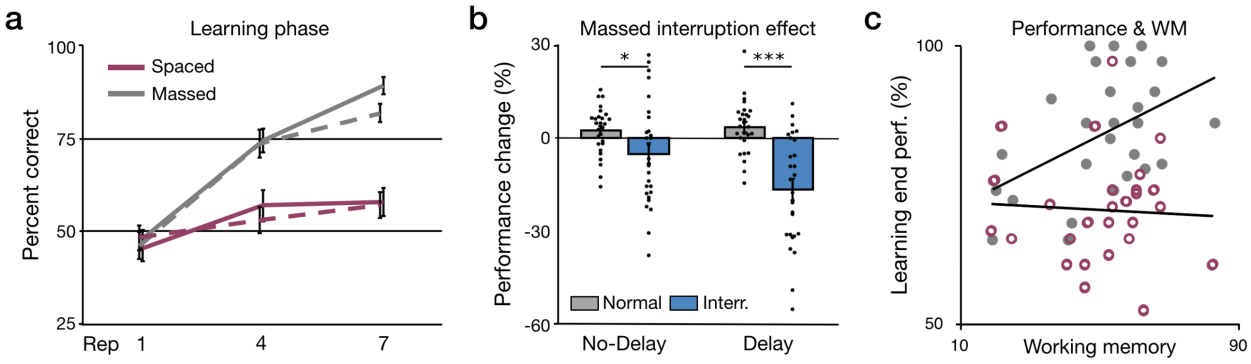

Figure 2. (**a**) Learning performance in the No-Delay group (dashed lines) and Delay group (solid lines) across the learning phase for massed-trained (grey) and spaced-trained (magenta) stimuli. (**b**) During learning, interruption of massed training by occasional sets of spaced trials was followed by a relative decrease in post-interruption performance in both the No-Delay and Delay group (blue), relative to the normal learning-related increase in performance in a balanced no-interruption repetition measure (grey). Black dots represent individual participants. (**c**) A working memory measure was collected in the No-Delay group (O-SPAN). Working memory capacity was positively correlated with massed but not spaced learning performance (difference $p = 0.004$; massed, filled grey circles; spaced, open magenta circles). (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$)

Relationship between massed learning performance and working memory

As a third test of the relationship between performance on massed-trained and working memory, we expected that overall massed condition learning performance would be related to individual differences in working memory capacity (Wimmer and Poldrack, 2018). In the No-Delay group, we collected a separate measure of working memory capacity (O-SPAN). We found that performance in the massed but not the spaced condition was significantly correlated with working memory ($r = 0.56$, $p = 0.0024$; spaced, $r = -0.20$, $p > 0.31$; **Fig. 2c**). Further, the performance-WM correlation in the massed condition was significantly stronger than that in the spaced condition ($z =$

2.86, p = 0.004). Thus, three separate behavioral measures support a role for short-term memory in rapid massed learning: First, learning for massed-trained stimuli was faster than learning for spaced-trained stimuli. Second, learning for massed-trained stimuli was negatively affected by interruptions by spaced stimulus trials. Third, massed learning performance was positively correlated with individual differences in working memory capacity.

Choice test. Next, we examined the question of whether a delay after learning affected performance in the test phase incentive-compatible choices. We focused on whether the delay led to relative decreases in performance for massed- versus spaced-trained stimuli. In the choice test phase, participants were presented with two stimuli from the learning phase and were asked to choose the stimulus with the higher value.

In the No-Delay group, choice performance overall was greater than chance for both massed- and spaced-trained stimuli (massed: 76.6 % CI [68.4 84.9]; $t_{(26)} = 6.64$, $p < 0.001$; spaced: 63.0 % CI [53.5 72.4]; $t_{(26)} = 2.82$, $p = 0.009$). Similar to performance at the end of the preceding learning phase, we found that test phase performance was higher for massed- vs. spaced-trained stimuli ($t_{(26)} = 2.81$, CI [3.7 23.7]; $p < 0.001$; **Fig. 3a**). We did not find an effect of learning order on massed condition choice test performance (including participants with below-ceiling choice performance (n = 20), β = 0.004 [-0.017 0.027]; t = 0.404, p = 0.695; TOST p = 0.010).

To measure the change from the end of the learning phase to the choice test phase, we computed the difference in performance across phases. In the No-Delay condition, we found that relative performance in the massed-trained condition tended to decrease and relative performance in the spaced-trained condition tended to increase,

although this interaction was not significant ($t_{(26)}$ = -1.83, CI [-23.7 1.4]; p = 0.078; TOST p = 0.135; **Fig. 3b**). It is difficult to interpret a change within either the massed or the spaced condition separately, given the different underlying performance measures during learning and test, but the numerical shift was similar and opposing for the massed and spaced conditions (massed: -5.4 % CI [-14.5 3.7]; $t_{(26)}$ = -1.22, p = 0.24; spaced: 5.8 % CI [-3.0 14.6]; $t_{(26)}$ = 1.36, p = 0.19).

In the Delay group, where the choice test started more than 15 minutes after the learning phase ended, choice performance was also greater than chance for both massed- and spaced-trained stimuli (massed: 75.4 % CI [66.3 84.5]; $t_{(26)}$ = 5.75, p < 0.001; spaced: 76.1 % CI [66.1 86.2]; $t_{(26)}$ = 5.34, p < 0.001). In strong contrast to the end of learning phase performance as well as performance in the No-Delay group, however, we found that choice performance was numerically matched for massed- and spaced-trained stimuli ($t_{(26)}$ = -0.13, CI [-11.6 10.3]; p = 0.90; TOST p = 0.004; **Fig. 3a**). We also found an effect of learning order on massed condition choice test performance, such that performance was higher for stimuli presented closer to the test phase (including participants with below-ceiling choice performance (n = 18), β = 0.036 [0.007 0.065]; t = 2.457, p = 0.013).

Critically, in the Delay group, the change in relative performance between the end of learning and test revealed a significant interaction with spacing condition ($t_{(26)}$ = -5.07, CI [-45.0 -19.0]; p < 0.001; **Fig. 3b**), supported by a relative decrease in performance across the rest period for massed-trained stimuli and a relative increase in performance for spaced-trained stimuli. Although changes from learning to test within the massed and spaced conditions are difficult to interpret separately (as noted above),

the effects were in the expected direction (massed: -13.8 % CI [-23.5 -4.1]; $t_{(26)}$ = -2.92, p = 0.007; spaced: 18.2 % CI [8.2 28.2]; $t_{(26)}$ = 3.73, p < 0.001).

Next, we compared the change in performance from learning to test between the Delay and No-Delay groups. We found that the interaction between the post-learning performance change and rest delay was significant ($t_{(52)}$ = 2.37, CI [3.2 38.4]; p = 0.022; **Fig. 3b**). This interaction was supported by a numerical performance decrease across the delay in the massed condition (-8.40 % CI [-21.4 4.6]; $t_{(52)}$ = -1.297, p = 0.20; TOST p = 0.214, indicating that we cannot rule out the presence of a medium-sized effect) and a positive but non-significant increase across the delay in the spaced condition (12.40 % CI [-0.6 25.4]; $t_{(52)}$ = 1.910, p = 0.062; TOST p = 0.154; **Fig. 3b**, magenta bars). Also supporting a delay-related increase in performance for spaced-trained stimuli, we found a positive but non-significant increase in overall choice performance between the No-Delay group and the Delay group ($t_{(26)}$ = -1.958 CI [-26.6 0.3]; p = 0.056; **Fig. 3a**, magenta bars).

These results indicate that rest has a detrimental effect on performance for massed-trained stimuli, but a positive effect on choice test performance for spaced-trained stimuli. Overall, the decrease in massed relative to spaced condition performance supports a role for a short-term working memory-like mechanism during learning that fails to maintain massed-trained value associations, while the relative increase in spaced condition performance supports a separate mechanism that may increase the strength of value associations during rest.

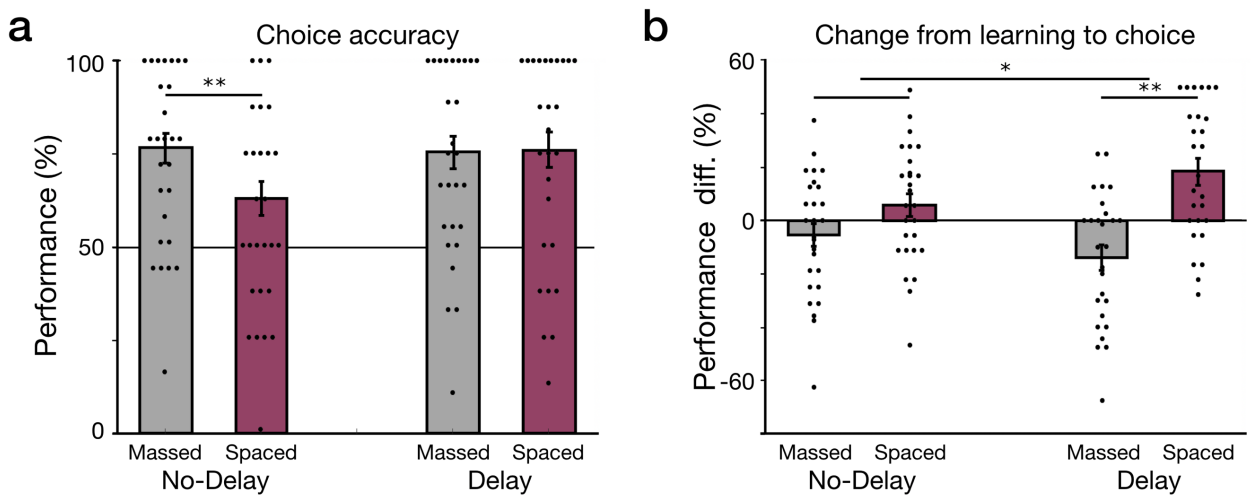

Figure 3. Choice test performance. (**a**) Choice test phase accuracy for the No-Delay and Delay groups (massed condition in grey; spaced condition in magenta. (**b**) Change from the end of learning to the choice test phase. In the Delay group, the relative performance change was significantly different, reflecting a decrease in performance for massed-trained stimuli and an increase in performance for spaced-trained stimuli. Across groups, the interaction of the performance difference with delay was significant. Black dots represent individual participants. (*$p < 0.05$; **$p < 0.01$)

We also collected choices between massed- and spaced-trained stimuli. In the No-Delay group, for reward-associated stimuli, we found that participants preferred massed-trained stimuli over the spaced-trained stimuli (massed versus spaced: 64.2 % CI [56.4 71.9]; $t_{(26)} = 3.767$, $p < 0.001$), consistent with the dominance of shorter-term massed-trained values on choices. For loss-associated stimuli, we found no difference in preference (51.5 % CI [39.0 58.0]; $t_{(26)} = 0.334$, $p = 0.74$; TOST $p = 0.007$). Reflecting the shift in preference across the delay, in the Delay group we found that participants showed no preference for massed- or spaced-trained stimuli (reward-associated massed versus spaced: 52.5 % CI [43.8 61.1]; $t_{(26)} = 0.588$, $p = 0.56$; TOST $p = 0.007$; loss-associated: 49.7 % CI [40.1 59.3]; $t_{(26)} = -0.066$, $p = 0.95$; TOST $p = 0.004$).

Comparing performance in these mixed choices across the No-Delay and Delay groups, we found that for reward-associated stimuli, preference for massed- versus spaced-trained stimuli was significantly lower in the Delay versus No-Delay group (reward difference: -11.7 % CI [-23.0 -0.4]; $t_{(52)} = -2.079$; $p = 0.043$; loss difference, -1.9 % CI [-15.0 11.3]; $t_{(52)} = -0.28$; $p = 0.779$; TOST $p = 0.011$). This relative decrease in value for massed and increase in value for spaced reward-associated stimuli after rest parallels the changes in performance from learning to test.

Post-learning reward ratings. Before the primary choice test measure, we also collected a supplemental reward association rating for each stimulus. Participants rated each stimulus on a graded scale, anchored by 0 % reward and 100 % reward. In the No-Delay group, graded reward association ratings were significantly higher for reward- versus loss- associated stimuli in the massed condition (reward: 65.5 % CI [60.0 70.9]; loss: 40.5 % CI [34.6 46.4]; difference $t_{(26)} = 5.66$, CI [15.9 34.0]; $p < 0.001$) and spaced condition (reward: 56.6 % CI [49.8 63.4]; loss: 35.0 % CI [28.7 41.4]; difference $t_{(26)} = 4.74$, CI [12.3 31.0]; $p < 0.001$). We did not find a difference in reward- versus loss-associated stimulus ratings based on spacing condition ($t_{(26)} = -0.18$, CI [-12.5 10.4]; $p = 0.858$; TOST $p = 0.014$). However, ratings were higher in the massed than in the spaced condition for both reward- and loss-associated stimuli ($t_{(26)} = 2.56$, CI [1.40 12.9]; $p = 0.017$).

In the Delay group, we also found that reward ratings were significantly greater for reward- vs. loss-associated stimuli in the massed condition (reward: 62.3 % CI [56.0 68.5]; loss: 39.6 % CI [33.6 45.6]; difference: $t_{(26)} = 5.16$, CI [13.6 31.7]; $p < 0.001$) as well as the spaced condition (reward: 56.7 % CI [49.1 64.4]; loss: 33.1 % CI [26.3 39.9];

difference: $t_{(26)}$ = 4.18, CI [12.0 35.3]; $p < 0.001$). We did not find a difference in reward- versus loss-associated stimulus ratings based on spacing ($t_{(26)}$ = -0.18, CI [-12.5 10.4]; p = 0.86; TOST p = 0.005). As in the No-Delay group, across both reward- and loss-associated stimuli, ratings in the massed condition were higher than in the spaced condition, although this did not reach significance ($t_{(26)}$ = 1.94, CI [-0.34 12.40]; p = 0.060).

We found no evidence that graded reward ratings differed due to spaced versus massed training in the No-Delay versus Delay groups ($t_{(52)}$ = 0.565, CI [-10.97 19.57]; p = 0.575; TOST p = 0.066). To further investigate this apparent difference between the choice test and the reward ratings results, we examined other potential effects of delay on reward ratings.

First, we examined whether ratings decreased in uncertainty over time, operationalized as variability. We found a tendency for an effect of delay on ratings variability, such that variability for spaced-trained stimuli was relatively lower after a delay. Specifically, in the No-Delay group, we found no difference in ratings variability between massed- and spaced-trained stimuli (massed mean of within-participants standard deviation: 23.0 CI [19.8 26.2]; spaced 22.4 CI [18.5 26.4]; difference: $t_{(26)}$ = 0.26, CI [-3.9 5.0]; p = 0.797; TOST p = 0.006). In the Delay group, however, we found that ratings variability for spaced-trained stimuli was significantly lower than variability for massed-trained stimuli (massed: 23.6 CI [20.6 26.5]; spaced: 18.3 CI [14.8 21.8]; difference: $t_{(26)}$ = -2.44, CI [-9.7 -0.8]; p = 0.022). This effect was not significantly different between the No-Delay versus the Delay group ($t_{(52)}$ = -1.543, CI [-10.8 1.4]; p = 0.129; TOST p = 0.084). These results ratings show that after a delay, value estimates

as reflected in ratings for spaced-trained stimuli were associated with relatively lower uncertainty.

Next, we examined the relationship between reward ratings and test phase choices. As expected, we found a positive relationship between post-learning ratings and choices. Ratings in the No-Delay group were significantly related to test phase choices (massed: β = 0.010 CI [0.008 0.011]; $t_{(26)}$ = 11.62, $p < 0.001$; spaced: β = 0.010 CI [0.006 0.012]; $t_{(26)}$ = 6.74, $p < 0.001$), with no difference based on condition ($t_{(26)}$ = 1.08, CI [-0.001 0.004]; p = 0.29; TOST p = 0.036). In the Delay group, ratings were also related to subsequent test phase choices (massed: β = 0.010 CI [0.007 0.012]; $t_{(26)}$ = 7.16, $p < 0.001$; spaced: β = 0.013 CI [0.011 0.015]; $t_{(26)}$ = 14.22, $p < 0.001$). Critically, in the Delay group, we found a significantly weaker relationship between ratings and choices in the massed compared to the spaced condition ($t_{(26)}$ = -2.37, CI [-0.007 -0.001]; p = 0.026).

Comparing the relationship between ratings and choices across the No-Delay and Delay groups, we found that a delay before the test led to a significant increase in the relationship between ratings and choices in the spaced versus the massed condition ($t_{(52)}$ = -2.52, CI [-0.009 -0.001]; p = 0.015). This difference between the No-Delay and Delay groups was also reflected within the spaced condition alone ($t_{(52)}$ = -2.57, CI [-0.007 -0.001]; p = 0.013). Thus, even though ratings appeared to show similar differentiation of reward- versus loss-associated stimuli across conditions in each group, we found that ratings for massed-trained stimuli were relatively worse predictors and ratings for spaced-trained stimuli were relatively better predictors of choices after a rest delay. This pattern is consistent with a decay in the fidelity of massed-trained value associations and an increase in the fidelity of spaced-trained value associations. While

the incentive-compatible choice test is our primary outcome measure, the underlying patterns in the reward rating data are consistent with opposing effects of a rest delay on the variability of learned values and choice performance for massed- versus spaced-trained stimuli.

Discussion

Reward learning in humans is likely to be supported by an interaction between working memory and reinforcement learning mechanisms (Collins and Frank, 2018; Wimmer et al., 2018). We examined the effect of massed versus spaced learning of reward and loss associations and the effect of post-learning rest on memory for value associations. The close repetition of trials in our “massed” condition was designed to be similar to many feedback-based learning designs used in human research (e.g. Daw et al., 2011; Wimmer et al., 2014). During learning, repetitions for massed stimuli were presented sequentially with occasional interruptions for the presentation of “spaced” stimuli. Across multiple measures and two experimental groups, we found that short-term working memory resources supported massed performance. First, we found that massed training led to better overall performance during the learning phase. Second, brief interruptions of massed training were followed by performance decreases. Third, an independent measure of working memory capacity was related to massed but not spaced learning performance. Finally, we found an effect of post-learning delay: when tested with no delay, performance was similar to performance at the end of learning. However, when tested after a delay, relative performance for massed-trained stimuli decreased while performance for spaced-trained stimuli increased.

Working memory and learning phase performance

Our results during the learning phase provide new support for an important role of short-term memory in typical reward-based and reinforcement learning paradigms. Recently, in conditions resembling massed learning, we found a positive correlation between working memory and learning performance (Wimmer et al., 2018). Previous

work supporting a negative interaction between episodic memory encoding and reinforcement learning suggested that this interaction may be due to competition over short-term memory processes and attention (Wimmer et al., 2014). Building on this work, the current paradigm provides both individual difference and trial-by-trial measures of the interaction between working memory and reward-based or reinforcement learning in the same task.

A closely related line of research on working memory and learning has indirectly investigated the effect of within-session spacing (Collins and Frank, 2012; Collins et al., 2014; Collins et al., 2017; Collins, 2018; Collins and Frank, 2018). In the paradigm used by Collins et al., participants engage in learning which of 3 response options are reinforced for a stimulus; across blocks, the number of concurrently learned stimuli within a block ("load") is systematically manipulated. Notably, by changing the number of stimuli concurrently presented within a block, this task also manipulates the spacing between repetitions of a given stimulus. While reinforcement learning models that only acquire stimulus-action values ("model-free" RL) predict that the number of stimuli in a block should have no effect on performance, strikingly, Collins et al. find that ongoing learning performance is significantly decreased as more stimuli are included in the learning block. This performance decrease with increasing load was successfully modeled using a reinforcement learning model that included a fixed-capacity working memory module capable of maintaining stimulus-response mappings (see also van de Vijver and Ligneul, 2019).

More recently, the effect of learning phase load on later choices was also investigated (Collins et al., 2017; Collins, 2018), with several similarities to the current paradigm's use of separate learning and choice test phases. In these studies, the test

phase followed a distractor task at a very similar delay to that used in our Delay group. The authors report relatively worse test phase performance for stimuli learned in low set-size blocks (similar to the current massed condition), suggesting that short-term working memory supports learning and then fails to support later test performance (Collins et al., 2017; Collins, 2018).

The current work extends these findings in several ways. We combine conditions favoring short-term over longer-term learning into the same relatively short learning session. We find that an independently collected measure of working memory capacity correlates with massed (but not spaced) learning performance (Wimmer and Poldrack, 2018). At test, we found that participants exhibited a preference for reward-associated stimuli from the massed versus spaced conditions in the No-Delay group, similar to a finding reported in Collins et al. (2017). However, by manipulating the delay period across groups, we show that performance for massed-trained stimuli was intact immediately, but that preference significantly decays after a delay. The change in performance from the end of learning to test phase choices was negative for massed-trained stimuli but positive for spaced-trained stimuli.

The negative trial-by-trial interruption effect and the negative post-learning delay effect on massed performance indicate different timecourses of forgetting for massed-trained associations. These findings suggest extensions or modifications to the single probabilistic working memory model proposed previously (Collins and Frank, 2012). Additionally, in this model the working memory module (with a 4-7 stimulus capacity) would be unable to maintain the 14 value representations needed to support performance in the No-Delay testing condition in the current experiment. Instead of a single separate working memory module with a single rate of forgetting, our current

results and related findings (Wimmer and Poldrack, 2018) indicate that a short-term memory mechanism with an exponentially decreasing forgetting rate, or multiple neural stores with graded degrees of decay would better capture the full range of observed behavior.

Effect of post-learning delay period

Critically, we found an effect of rest delay on memory for value associations, where relative performance decreased for massed-trained stimuli and increased for spaced-trained stimuli. Our results suggest that a post-learning delay period may both impair learning for massed-trained value associations and improve learning for spaced-trained value associations (Schmidt and Bjork, 1992; Taylor and Rohrer, 2010). However, we cannot directly attribute these changes to effects of either the massed or spaced training alone. The direct comparisons of performance from the No-Delay to the Delay groups for massed- or spaced-trained stimuli were not significant, although our results indicate that we cannot rule out the presence of a medium-sized effect in each condition (especially for spaced-trained stimuli).

For massed-trained items, a straightforward prediction is that associations that are dependent on short-term memory will decay over time. If short-term memory resources have an exponentially decreasing forgetting rate, this predicts a decline in performance across longer rest periods. For spaced-trained items, the change in performance at test cannot be explained as a simple effect of impaired performance during learning, as learning phase performance was similar in both the No-Delay and Delay groups. One possibility is that value associations that are weakly encoded during the learning phase are “sharpened” during the post-learning delay period. This

interpretation is further supported by the significant decrease in post-learning reward rating variability after a delay.

Previous research on the benefits of spacing has also found evidence of a positive effect of delay between learning and test in verbal memory and motor domains (Lee and Genovese, 1988; Donovan and Radosevich, 1999; Janiszewski et al., 2003; Cepeda et al., 2006). In this work, performance has been shown to be relatively impaired by spaced training while later retention is improved (Schmidt and Bjork, 1992; Taylor and Rohrer, 2010; Soderstrom and Bjork, 2015). To our knowledge, a benefit of a post-learning delay period has not been demonstrated for learning simple stimulus-reward associations in humans. However, a previous study that employed a more elaborate multi-step associative structure found that post-learning rest led to improved test performance (Gershman et al., 2014).

Positive effects of a delay period could come from spontaneous reactivation of spaced-trained associations after learning (Gomperts et al., 2015; Gruber et al., 2016; Olafsdottir et al., 2018), a mechanism implemented in the DYNA reinforcement learning model (Sutton, 1990). Moreover, spacing may also lead to more lasting learning due to the unexpected nature of spaced item appearance during training (Bouton and Moody, 2004). Relative novelty has been demonstrated to interact with reward processing, leading to increased activity in the hippocampus as well as the striatum (Guitart-Masip et al., 2010; Bunzeck et al., 2011; Zaehle et al., 2013). This effect may be driven by a hippocampal-VTA loop, supporting an increase in the firing rate of dopamine neurons in response to rewards for spaced stimuli (Lisman and Grace, 2005; Goto and Grace, 2008). Finally, spacing of learning repetitions may also allow for relatively short-term

synaptic plasticity mechanisms to iteratively build stronger associations (Reynolds et al., 2001).

It is possible that a longer delay before testing would reveal more robust changes in performance at test. Supporting this view, we previously found that after 3 weeks, associations that had been trained over multiple sessions were robust to forgetting, unlike associations trained in a single session (Wimmer et al., 2018). Similar effects were recently reported in a related study (van de Vijver and Ligneul, 2019). Notably, however, our current results indicate that spacing can be successfully studied in just a single session.

Conclusion

In summary, we found that reward-based learning with typical massed presentation of stimuli shows a strong dependence on short-term working memory resources across multiple measures. Further, after learning, a delay period led to a relative decrease in performance for massed-trained stimuli and increase in performance for spaced-trained stimuli. These and other related findings indicate that studies of human reward-based learning that employ common closely-spaced learning designs are likely a suboptimal way to measure individual differences in the mechanisms supporting reward-based or reinforcement learning (Collins and Frank, 2012; Collins et al., 2014; Collins, 2018; Wimmer et al., 2018). A classic distinction between short-term performance versus later evidence of learning is helpful to remember when examining the utility of common reward-based learning paradigms (Schmidt and Bjork, 1992; Taylor and Rohrer, 2010; Soderstrom and Bjork, 2015). As demonstrated across decades of research in motor and verbal learning, higher ongoing

performance can be supported by short-term mechanisms and is often not predictive of subsequent measures of learning. Importantly, we also replicate the finding that conditions that decrease short-term performance can actually lead to improvements in later learning measures (Schmidt and Bjork, 1992; Soderstrom and Bjork, 2015). In this light, many reinforcement learning studies in humans may be predominantly studying short-term reinforcement *performance*, not learning.

We recommend employing alternative experimental designs with increased spacing between training repetitions, as in the current spaced-training condition, which strongly limit the use of working memory and likely increase reliance on reward-based learning mechanisms in the brain (such as the striatal dopamine system). Additionally, spacing training across days can produce value associations that are resistant to forgetting (Kim et al., 2015; Wimmer et al., 2018), potentially due to better isolation of the neural mechanisms supporting reinforcement learning. Overall, learning in spaced designs may be more supported by the kind of mechanisms supporting reward-based learning in environments outside the lab, where learning repetitions are likely to be spread over periods of time longer than several seconds. Translationally, spaced training also shows promise for increasing our understanding of the characteristics of maladaptive value associations, such as those found in addiction (Kim et al., 2015; Wimmer et al., 2018), with further implications for research on reward-based learning in learning deficits in psychiatric disorders (Maia and Frank, 2011; Montague et al., 2012; Whitton et al., 2015; Huys et al., 2016; Moutoussis et al., 2016).

References

Bouton ME, Moody EW (2004) Memory processes in classical conditioning. Neuroscience and biobehavioral reviews 28:663-674.

Braun EK, Wimmer GE, Shohamy D (2018) Retroactive and graded prioritization of memory by reward. Nat Commun 9:4886.

Bunzeck N, Guitart-Masip M, Dolan RJ, Duzel E (2011) Contextual novelty modulates the neural dynamics of reward anticipation. J Neurosci 31:12816-12822.

Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D (2006) Distributed practice in verbal recall tasks: A review and quantitative synthesis. Psychol Bull 132:354-380.

Collins AG, Frank MJ (2012) How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. Eur J Neurosci 35:1024-1035.

Collins AG, Brown JK, Gold JM, Waltz JA, Frank MJ (2014) Working memory contributions to reinforcement learning impairments in schizophrenia. J Neurosci 34:13747-13756.

Collins AGE (2018) The tortoise and the hare: Interactions between reinforcement learning and working memory. J Cogn Neurosci 30:1422-1432.

Collins AGE, Frank MJ (2018) Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. Proc Natl Acad Sci U S A 115:2502-2507.

Collins AGE, Albrecht MA, Waltz JA, Gold JM, Frank MJ (2017) Interactions Among Working Memory, Reinforcement Learning, and Effort in Value-Based Choice: A New Paradigm and Selective Deficits in Schizophrenia. Biol Psychiatry 82:431-439.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204-1215.

Donovan JJ, Radosevich DJ (1999) A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. J Appl Psychol 84:795-805.

Ebbinghaus H (1913) Memory: A contribution to experimental psychology.

Gerraty RT, Davidow JY, Wimmer GE, Kahn I, Shohamy D (2014) Transfer of learning relates to intrinsic connectivity between hippocampus, ventromedial prefrontal cortex, and large-scale networks. J Neurosci 34:11297-11303.

Gershman SJ, Markman AB, Otto AR (2014) Retrospective revaluation in sequential decision making: a tale of two systems. J Exp Psychol Gen 143:182-194.

Gomperts SN, Kloosterman F, Wilson MA (2015) VTA neurons coordinate with the hippocampal reactivation of spatial experience. Elife 4.

Goto Y, Grace AA (2008) Dopamine modulation of hippocampal-prefrontal cortical interaction drives memory-guided behavior. Cereb Cortex 18:1407-1414.

Gruber MJ, Ritchey M, Wang SF, Doss MK, Ranganath C (2016) Post-learning Hippocampal Dynamics Promote Preferential Retention of Rewarding Events. Neuron 89:1110-1120.

Guitart-Masip M, Bunzeck N, Stephan KE, Dolan RJ, Duzel E (2010) Contextual novelty changes reward representations in the striatum. J Neurosci 30:1721-1726.

Huys QJ, Maia TV, Frank MJ (2016) Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci 19:404-413.

Janiszewski C, Noel H, Sawyer AG (2003) A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. J Consum Res 30:138-149.

Kim HF, Ghazizadeh A, Hikosaka O (2015) Dopamine Neurons Encoding Long-Term Memory of Object Value for Habitual Behavior. Cell 163:1165-1175.

Lakens D (2017) Equivalence tests: A practical primer for t-tests, correlations, and metaanalyses. Soc Psychol Personal Sci.

Lee TD, Genovese ED (1988) Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. Res Q Exercise Sport 59:277-287.

Lewandowsky S, Oberauer K, Yang LX, Ecker UK (2010) A working memory test battery for MATLAB. Behav Res Methods 42:571-585.

Lisman JE, Grace AA (2005) The hippocampal-VTA loop: controlling the entry of information into long-term memory. Neuron 46:703-713.

Maia TV, Frank MJ (2011) From reinforcement learning models to psychiatric and neurological disorders. Nat Neurosci 14:154-162.

Montague PR, Dolan RJ, Friston KJ, Dayan P (2012) Computational psychiatry. Trends Cogn Sci 16:72-80.

Moutoussis M, Eldar E, Dolan RJ (2016) Building a New Field of Computational Psychiatry. Biol Psychiatry.

Olafsdottir HF, Bush D, Barry C (2018) The Role of Hippocampal Replay in Memory and Planning. Current biology : CB 28:R37-R50.

Otto AR, Raio CM, Chiang A, Phelps EA, Daw ND (2013) Working-memory capacity protects model-based learning from stress. Proc Natl Acad Sci U S A 110:20941-20946.

Reynolds JN, Hyland BI, Wickens JR (2001) A cellular mechanism of reward-related learning. Nature 413:67-70.

Schmidt RA, Bjork RA (1992) New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. Psychological Science 3:207-217.

Schuirmann DJ (1987) A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm 15:657-680.

Soderstrom NC, Bjork RA (2015) Learning Versus Performance: An Integrative Review. Perspectives on Psychological Science 10:176-199.

Sutton RS (1990) Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In: Proceedings of the Seventh International Conference on Machine Learning (Porter BW, Mooney RJ, eds), pp 216-224: Morgan Kaufmann.

Taylor K, Rohrer D (2010) The effects of interleaved practice. Appl Cognitive Psych 24:837-848.

Teichner WH (1952) Experimental extinction as a function of the intertrial intervals during conditioning and extinction. Journal of Experimental Psychology 44:170-178.

Terrace HS, Gibbon J, Farrell L, Baldock MD (1975) Temporal factors influencing acquisition and maintenance of an autoshaped keypeck. Animal Learning & Behavior 3:53-62.

van de Vijver I, Ligneul R (2019) Relevance of working memory for reinforcement learning in older adults varies with timescale of learning. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn:1-23.

Whitton AE, Treadway MT, Pizzagalli DA (2015) Reward processing dysfunction in major depression, bipolar disorder and schizophrenia. Curr Opin Psychiatry 28:7-12.

Wimmer GE, Poldrack R (2018) Reinforcement learning over weeks versus minutes changes neural representations of value and leads to longer-lasting associations. biorXiv.

Wimmer GE, Braun EK, Daw ND, Shohamy D (2014) Episodic memory encoding interferes with reward learning and decreases striatal prediction errors. J Neurosci 34:14901-14912.

Wimmer GE, Li JK, Gorgolewski KJ, Poldrack RA (2018) Reward learning over weeks versus minutes increases the neural representation of value in the human brain. J Neurosci 38:7649-7666.

Yin HH, Knowlton BJ (2006) The role of the basal ganglia in habit formation. Nat Rev Neurosci 7:464-476.

Zaehle T, Bauch EM, Hinrichs H, Schmitt FC, Voges J, Heinze HJ, Bunzeck N (2013) Nucleus accumbens activity dissociates different forms of salience: evidence from human intracranial recordings. J Neurosci 33:8764-8771.