Reinforcement learning over time: spaced versus massed training establishes stronger value associations

Abbreviated title: Long-term reward learning

G. Elliott Wimmer and Russell A. Poldrack

Department of Psychology, Stanford University, 450 Serra Mall, Stanford, CA, 94305

Corresponding author:
G. Elliott Wimmer
Department of Psychology
Stanford University
450 Serra Mall, Bldg. 420 Jordan Hall
elliott@caa.columbia.edu

Acknowledgments

The authors thank Jamie Li for assistance in piloting, data acquisition, and interpretation, the help of Patrick Bissett during scanning, and the assistance of Ross Blair for the online testing platform. Research was supported by a research fellowship from the Deutsche Forschungsgemeinschaft and a pilot seed grant from the Stanford Center for Cognitive and Neurobiological Imaging to GEW. RAP is supported by the Laura and John Arnold Foundation and NIDA (UH2DA041713).

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 midbrain dopamine system. Learning paradigms are increasingly being extended to understand learning dysfunctions in mood and psychiatric disorders as well as addiction in the area of computational psychiatry. However, one potentially critical characteristic that this research ignores is the effect of time on learning: human feedback learning paradigms are conducted in a single rapidly paced session, while learning experiences in ecologically relevant circumstances and in animal research are almost always separated by longer periods of time. Event spacing is known to have strong positive effects on item memory across species and in reward learning in animals. Remarkably, the effect of spaced training on human reinforcement learning has not been investigated. In our experiments, we examined reward learning distributed across weeks vs. learning completed in a traditionally-paced or “massed” single session. Participants learned to make the best response for landscape stimuli that were either associated with a positive or negative value. In our first study, as expected, we found that after equal amounts of extensive training, accuracy was high and equivalent between the spaced and massed conditions. However, in a final online test 3 weeks later, we found that participants exhibited significantly greater memory for the value of spaced-trained stimuli. In our second study, our methods allowed for a direct comparison of maintenance of conditioning. We found that spaced training again had a beneficial effect: more than 87% of conditioning was maintained for spaced-trained stimuli, while only 30% was maintained for massed-trained stimuli. In addition, supporting a role for working memory in massed learning, across both studies we found a significant positive correlation between initial learning and working memory capacity. Our results indicate that single-session learning tasks may not lead to the kind of robust and lasting value associations that are characteristic of “habitual” value associations. Overall, these studies begin to address a large gap in our knowledge of fundamental processes of human reinforcement learning, with potentially broad implications for our understanding of learning in mood disorders and addiction.

Introduction

Rewarding and aversive experiences exert a strong influence on later decision making. When making a choice between an apple and a banana, for example, our decision likely relies on values shaped by countless previous experiences. In general, by learning over time which stimuli and actions regularly lead to favorable outcomes, we can make more adaptive choices in the future. Over the past few decades, neuroscience research has revealed the neural mechanisms supporting this kind of learning from reward feedback, demonstrating a critical role for the striatum and the midbrain dopamine system (Schultz et al., 1997; Rangel et al., 2008; Steinberg et al., 2013). Phasic activity in striatum-projecting dopamine neurons closely matches learning signals derived from reinforcement learning models in machine learning (Barto, 1995; Houk et al., 1995; Sutton and Barto, 1998).

More recently, researchers have also begun to apply reinforcement learning models to increase our understanding of behavioral dysfunctions in mood and psychiatric disorders as well as addictions in the growing area of “computational psychiatry” (Maia and Frank, 2011; Schultz, 2011; Montague et al., 2012; Whitton et al., 2015; Moutoussis et al., 2016). Foundational assumptions of this translational work on human reward-based learning are that learning behavior in these experiments is 1) supported by the same learning mechanisms illuminated in animal research, and 2) these learning mechanisms also support learning outside the lab. However, one important difference between feedback learning in the lab and learning in the real world is the effect of time: while the development of our habits and reward associations can involve separate learning events spread over days, weeks, and even years, human feedback learning paradigms involve only a single session, with events separated by seconds at most. At this condensed timescale, such “massed” learning paradigms allow processes other than habit learning, such as working memory, to dominate behavior (Collins and Frank, 2012; Collins et al., 2014).

The massed single-session nature of human learning experiments is in contrast to most animal experiments, where the neural systems supporting feedback learning were originally identified. In these experiments, learning spans hours and days in rodent studies or months in non-human primate studies (Schultz et al., 1997; Roesch et al., 2007). While functional imaging studies in humans have revealed robust correlates of

reward learning variables in human brain activity, in particular correlates of reward prediction error in the ventral striatum (Garrison et al., 2013), it is possible that the use of condensed learning sessions may lead to a distorted understanding of the underlying system which support reward learning. Thus, current feedback learning studies may not be able to provide a clear or complete understanding of the cognitive and neural mechanisms supporting gradual reward learning outside of the lab.

Previous research has shown powerful effects of spacing in various domains. Early research on memory for lists of items by Ebbinghaus (reported in Ebbinghaus, 1913) demonstrated a powerful beneficial effect of spaced learning on memory. Since these studies, research has continued to explore the positive effects of temporal spacing on episodic memory as well as learning in educational settings (Cepeda et al., 2006; Ellenbogen et al., 2007; Litman and Davachi, 2008; Carpenter et al., 2012). However, this kind of episodic or “relational” learning is known to depend on the medial temporal lobe and hippocampus, a separate, dissociable learning system from the striatal learning system (Eichenbaum and Cohen, 2001) (Knowlton et al., 1996). Importantly, early research on feedback learning also demonstrated a beneficial effect of spacing between learning events. A positive influence of spacing between one trial and another in simple conditioning has been shown in species ranging from Aplysia to honeybees to rats (Teichner, 1952; Carew et al., 1972; Terrace et al., 1975; Menzel et al., 2001). In these studies, learning is more rapid when events are spaced in time, even when there are fewer total learning events. Animal studies also hint at the possibility that feedback learning in spaced conditions is more robust, as demonstrated by less sensitivity to reward omission (Teichner, 1952). Surprisingly, only a handful of studies have examined the effect of spacing on human feedback learning, and all of these studies used a passive aversive learning task, eyeblink conditioning, which relies on a specialized cerebellar circuit (Humphreys, 1940; Spence and Norris, 1950; Kim and Thompson, 1997). Thus, very little is known about human feedback learning when event spacing more closely approximates how learning might occur in more ecologically relevant circumstances.

Our aim was to characterize the cognitive mechanisms which support learning long-term reward associations by spacing learning across days. We utilized a simple reward-based learning task where participants initially learned value associations for

spaced stimuli first in the lab and then online across two weeks using a novel online experiment portal (expfactory.org; Sochat et al., 2016). Associations for massed stimuli were only learned in a second in-lab session, closely matching the kind of training commonly used in reinforcement learning tasks. Finally, in order to study the maintenance of learning over longer timespans, we administered a delayed test three weeks after the completion of training.

Methods

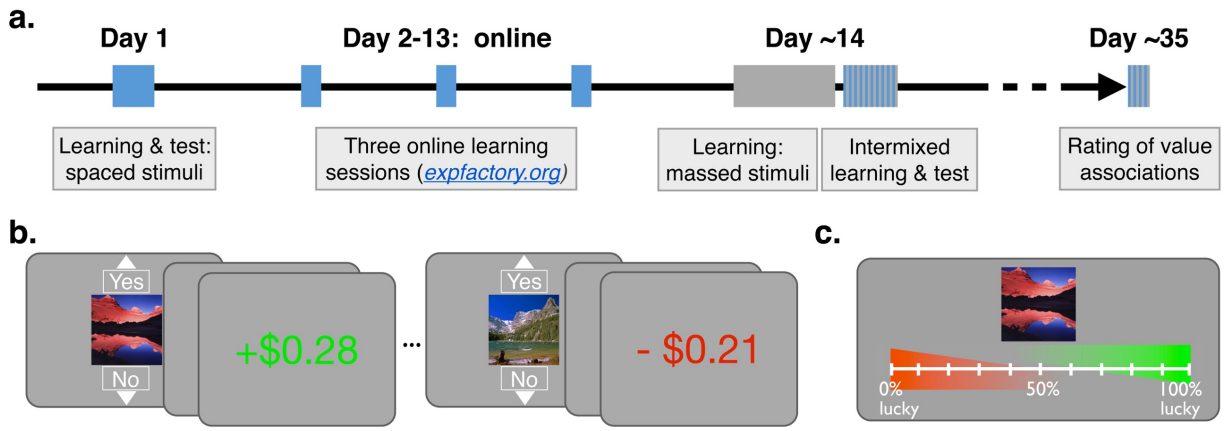

Figure 1. a) Experimental timeline. Learning for the spaced-trained stimuli is indicated in blue and learning for the massed-trained stimuli is indicated in grey. b) Reward learning task. Participants learned to select "Yes" for reward-associated stimuli and select "No" for loss-associated stimuli. c) Reward association rating test. This rating scale followed the initial in-lab learning sessions and was also administered 3 weeks after the last learning session.

Participants and Overview. Participants were recruited via advertising on the Stanford Department of Psychology paid participant pool web portal (https://stanfordpsychpaid.sona-systems.com). Informed consent was obtained in a manner approved by the Stanford University Institutional Review Board. In study 1, behavioral and fMRI data acquisition proceeded until fMRI seed grant funding was expired, leading to a total of 34 scanned participants in the reward learning task. In order to ensure that the fMRI sessions two weeks after the first in-lab session were fully subscribed, a total of 62 participants completed the first behavioral session. Of this group, a total of 28 participants did not complete the fMRI and behavioral experiment described below. The results of 33 participants (20 female) are included in the analyses and results, with a mean age of 22.9 years (range: 18-34). Of the initial 34 scanned participants, we excluded one participant who exhibited extreme motion during the fMRI scan, indicating an inability to follow instructions. Of the 33 included participants, in two participants, rating and choice phase data from the second in-lab session were lost due

to errors in data recording; behavioral data from all other session are included. Participants were paid $10/hour for the first in-lab session and $30/hour for the second in-lab (fMRI) session, plus monetary rewards from the learning phase and choice test phase. The current report focuses only on the behavioral results of Study 1. fMRI results will be reported separately.

In Study 2, a total of 35 participants participated in the first session of the experiment, but 4 were excluded from the final dataset, as described below. Our sample size was designed to approximately match the size of Study 1. The final dataset included 31 participants (24 female), with a mean age of 23.3 years (range: 18-32). Two participants failed to complete the second in-lab session and all data were excluded; one other participant exhibited poor performance the first session (less than 54% correct during learning and less than 40% correct in the choice test) and was therefore excluded from participation in the follow-up sessions. Of the 31 included participants, one participant failed to complete the third in-lab session, but data from other sessions was included. Participants were paid $10/hour for the two in-lab sessions, monetary rewards from the learning phase and choice test phase, plus a bonus of $12 for the 5-minute duration third in-lab session.

Both Study 1 and Study 2 utilized the same reward-based learning task. Participants learned the best response for individual stimuli in order to maximize their payoff. Two different sets of stimuli were either trained across two weeks (“spaced-trained” stimuli) or in a single session (“massed-trained” stimuli; **Figure 1a**). Spaced training began in the first in-lab session and continued across three online training sessions spread across approximately 2 weeks. Training on massed stimuli began in the second in-lab session. Spaced training always preceded massed training, so that by the end of the second in-lab session both sets of stimuli had been shown on an equal number of learning trials. This design was the same across Study 1 and Study 2, with the difference that Study 1 included an fMRI portion. Additionally, the three-week follow-up measurement was conducted online for Study 1 and in-lab for Study 2.

Experimental design, Study 1. In Study 1, before the learning phase, participants rated a set of 38 landscape picture stimuli based on liking, using a computer mouse, preceded by one practice trial. The same selection procedure and landscape stimuli

were used previously (Wimmer and Shohamy, 2012). These ratings were used to select the 16 most neutrally-rated set of stimuli per participant to be used in Study 1. Stimuli were then randomly assigned to condition (spaced or massed) and value (reward or non-reward). In Study 2, we used the ratings collected across participants in Study 1 to find the most neutrally-rated stimuli on average and then created two counterbalanced lists of stimuli from this set.

Next, in the reward game across both studies, participants learned the best response (arbitrarily labeled “Yes” and “No”) for each stimulus. Participants used up and down arrow keys to make “Yes” and “No” responses, respectively. Reward-associated stimuli led to a win of $0.35 on average when “Yes” was selected and a small loss of -$0.05 when “No” was selected. Non-reward-associated stimuli led to a neutral outcome of $0.00 when “No” was selected and -$0.25 when “Yes” was selected. These associations were probabilistic, such that the best response led to the best outcome 80% of the time during training. If no response was recorded, at feedback a warning was given: “Too late or wrong key! - $0.50”, and participants lost $0.50.

To increase engagement and attention to the feedback, we introduced uncertainty into the feedback amounts in two ways: first, all feedback amounts were jittered ± $0.05 around the mean using a flat distribution. Second, for the reward-associated stimuli, half were associated with a low reward amount ($0.45) and half with a higher reward amount ($0.25). We did not find that this second manipulation significantly affected learning performance at the end of the training phase, and thus our analyses and results collapse across the reward levels.

In a single reward learning trial, a stimulus was first presented with the options “Yes” and “No” above and below the image, respectively (**Figure 1b**). Participants had 2 seconds to make a choice. After the full 2 s choice period, a 1 s blank screen ITI preceded feedback presentation. Feedback was presented in text for X s, leading to a total trial duration of 1.5 s. Reward feedback above +$0.10 was presented in green, and feedback below $0.00 was presented in red, while other values were presented in white. After the feedback, an ITI of duration 2 preceded the next trial (min, 0.50 s; max, 3.5 s), where in the last 0.25 s prior to the next trial the fixation cross turned from white to black. The background for all parts of the experiment was grey (RGB value [111 111 111]).

In the first in-lab session, participants learned associations for spaced-trained stimuli, which differed from the training for massed-trained stimuli only in that training for spaced stimuli was spread across 4 sessions, 1 in-lab and 3 online. Training for massed-trained stimuli only occurred in the subsequent second in-lab session. Initial learning for both spaced-and massed-trained stimuli included 8 stimuli, of which half were rewarded and half were non-rewarded. In the initial learning phase for both conditions, each stimulus was repeated 10 times. The lists for the initial learning session were pseudo-randomized, with constraints introduced to facilitate initial learning and to achieve ceiling performance before the end of training. In the first learning session for both spaced- and massed-trained stimuli, 4 stimuli were introduced in the first 40 trials and the other 4 stimuli were introduced in the second 40 trials. Further, when a new stimulus was introduced, the first repetition followed immediately. The phase began with 4 practice trials including 1 reward-associated practice stimulus and 1 non-reward-associated practice stimulus, followed by a question about task understanding. Three rest breaks were distributed throughout the rest of the phase.

After the initial learning phase in both conditions, participants completed a rating phase and a choice phase. In the rating phase, participants tried to remember whether a stimulus was associated with reward or not. They were instructed to use a rating scale to indicate their memory and their confidence in their memory using a graded scale, with responses made via computer mouse (**Figure 1c**). Responses near the scale line were recorded. Responses were self-paced. After 0.5 s, trials were followed by a 3 s ITI. For analyses, responses (recorded in pixel left-right location values) were transformed to 0-100 percent.

In the choice phase, participants made a forced-choice response between two stimuli, only including spaced stimuli in the first in-lab session and only including massed stimuli in the second in-lab session. Stimuli were randomly presented on the left and right side of the screen. Participants made their choice using the 1-4 number keys in the top row of the keyboard, with a ‘1’ or ‘4’ response indicating a confident choice of the left or right option, respectively, and a ‘2’ or ‘3’ response indicating a guess choice of the left or right option, respectively. The trial terminated 0.25 s after a response was recorded, followed by a 2.5 s ITI. Responses were self-paced. Participants were informed that they would not receive feedback after each choice but

that the computer would keep track of the number of correct choices of the reward-associated stimuli that were made and pay a bonus based on their performance. As the long-term follow-up only included ratings, choice analyses were limited to comparing how choices aligned with ratings.

At the end of the session, participants completed the Beck Depression Inventory (BDI) and the operation-span task (OSPAN) to collect a measure of individual working memory capacity (Lewandowsky et al., 2010; Otto et al., 2013). In the operation-span task, 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. After arithmetic-letter sequences ranging in length from 4 to 8, participants were asked to type in the letters that they had seen in order, with no time limit. Each sequence length was repeated 3 times. 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, 49.9 ± 3.0; range, 19-83) (Otto et al., 2013).

Subsequent to the first in-lab session where training on spaced stimuli began, participants completed three online sessions with the spaced-trained stimuli. Sessions were completed on a laptop or desktop computer (but not on mobile devices), using the expfactory.org platform (Sochat et al., 2016). Code for the online reward learning phase can be found at: https://github.com/gewimmer/reward_learning. Each online training session included 5 repetitions of the 8 spaced-trained stimuli, in a random order, leading to 15 additional repetitions per spaced-trained stimulus overall. The task and timing was the same as in the in-lab sessions, with the exception that the screen background was white and white feedback text was replaced with grey. Participants completed the online sessions across approximately 2 weeks, initiated with an email from the experimenter including login details for that session. In the case that participants had not yet completed the preceding online session when the notification about the next session was received, participants were instructed to complete the preceding session that day and the next session the following day. Thus, at least one overnight period was required between sessions. Participants were instructed to complete the session when they were alert and not distracted.

Next, participants returned for a second in-lab session, approximately two weeks later (mean, 13.6 days; range, 10-20 days). Here, participants began and completed training on the massed-trained stimuli. Initial training across the first 10 repetitions was conducted as described above for the first in-lab session. Next, participants completed a rating phase including both spaced- and massed-trained stimuli and choice phase involving only the massed-trained stimuli. After this, participants finished training on the massed-trained stimuli, bringing total experience up to 25 repetitions, the same as for the spaced-trained stimuli to that point.

In Study 1, participants next entered the scanner for an intermixed training session. Across 3 blocks, participants engaged in additional training on the spaced- and massed-trained stimuli, with 6 repetitions per stimulus. During scanning, ITI durations were on average 3.5 s (min, 1.45 s; max, 6.55 s). Responses were made using a button cylinder, with the response box positioned to allow finger responses to mirror those made on the up and down arrow keys on the keyboard. Following the intermixed training session, participants engaged in a no-feedback block, where stimuli were presented with no response requirements. To provide a measure of attention and to promote recollection and processing of stimulus value, participants were instructed to remember whether a stimulus had been associated with reward or with no reward. On ~10% of trials, after the stimulus had disappeared, participants were asked to answer whether the best response to the stimulus was a “Yes” or a “No”. Each stimulus was repeated 10 times during this no-feedback phase.

After scanning, participants engaged in an exploratory block to study whether and how participants would reverse their behavior given a shift in feedback contingencies. Importantly, the “reversed” stimuli (2 per condition per participant) were not included in the analyses of the 3-week follow-up data. In brief, we did not find any differences between spaced- and massed-trained stimuli during reversal or in post-reversal ratings, and thus these results are not discussed further. For further methods and results of this phase, please see Supplementary Information.

We administered a follow-up test of memory for the value of conditioned stimuli approximately 3 weeks later (mean, 24.5 days; range, 20-37 days). An online questionnaire was constructed with each participant’s stimuli using Google Forms (https://docs.google.com/forms). Participants were instructed to try to remember

whether a stimulus was associated with winning money or not winning money, using an adapted version of the scan from the rating phase of the in-lab experiment. Responses were recorded using a 10-point radio button scale, anchored with “0% lucky” on the left to “100% lucky” on the right. Similar to the in-lab ratings, participants were instructed to respond to the far right end of the scale if they were completely confident that a given stimulus was associated with reward and to the far left if they were completely confident that a given stimulus was associated with no reward. Thus, distance from the center origin represented confidence in their memory.

In-lab portions of the study were presented using Psychtoolbox 3.0 (Brainard, 1997), with the initial in-lab session conducted on 21.5” Apple iMacs. Online training was completed using expfactory.org (Sochat et al., 2016), with functions adapted from the jspsych library (de Leeuw, 2015). At the second in-lab session, before scanning, participants completed massed-stimulus training on a 15” MacBook Pro laptop. During scanning, stimuli were presented on a screen positioned above the participant’s eyes that reflected an LCD screen placed in the rear of the magnet bore. Responses during the fMRI portion were made using a 5-button cylinder button response box (Current Designs, Inc.). Participants used the top button on the side of the cylinder for “Yes” responses and the next lower button for “No” responses. We positioned the response box in the participant’s hand so that the arrangement mirrored the relative position of the up and down arrow keys on the keyboard from the training task sessions.

Experimental design, Study 2. The methods for Study 2 were the same as in Study 1, with three main differences: training for massed stimuli was conducted without interruption for intermediate ratings, fMRI data were not collected, and the long-term follow-up was conducted in the lab rather than online.

Stimuli for Study 2 were composed of the most neutrally-rated stimuli from Study 1 pre-experiment ratings. Two counterbalance stimulus lists were created and assigned randomly to participants. The initial learning session for the spaced-trained stimuli and the three online training sessions were completed as described above. Following the training and testing phases, participants completed the OSPAN to collect a measure of working memory capacity. Scores were calculated as in Study 1 (mean, 49.7 ± 3.2; range 17-83).

The second in-lab session was completed approximately two weeks after the first session (mean, 12.8 days; range, 10-17 days). Here, participants learned reward associations for the set of "massed" stimuli. The training progressed through all 25 repetitions of the massed-trained stimuli with only short rest breaks, omitting the intervening test phases of Study 1. In the last part of the learning phase, to assess end-state performance on both spaced-trained and massed-trained stimuli, 3 repetitions of each stimulus were presented in a pseudo-random order. Rating and choice phase data were acquired after this learning block, with trial timing as described above.

After the choice phase, we administered an exploratory phase to assess potential conditioned stimulus-cued biases in new learning. This phase was conducted in a subset of 25 participants, as the task was still under development when the data from the initial 6 participants were acquired. In brief, participants engaged in learning about new stimuli (abstract characters) in the same paradigm as described above (**Figure 1b**) while unrelated spaced- or massed-trained landscape stimuli were presented in the background during the choice period. We tested for but did not find any influence. For further methods and results, please see Supplementary Information.

Approximately 3 weeks after the second in-lab session (mean, 21.1 days; range, 16-26 days), participants returned to the lab for the third and final in-lab session. Using the same testing rooms as during the previous sessions (which included the full training session on massed stimuli), participants completed another rating phase. Participants were reminded of the reward rating instructions and told to "do their best" to remember whether individual stimuli had been associated with reward or non-reward during training. Trial timing was as described above, and the order of stimuli was pseudo-randomized.

Analysis. Behavioral analyses were conducted in Matlab 2016a (The MathWorks, Inc., Natick, MA). Results presented below are from the following analyses: t-tests vs. chance for learning performance, within-group (paired) t-tests comparing differences in reward- and non-reward-associated stimuli across conditions, Pearson correlations, and Fisher z-transformations of correlation values. We additionally tested whether non-significant results were weaker than a moderate effect size using the Two One-Sided Test (TOST) procedure (Schuirmann, 1987; Lakens, 2017) and the TOSTER library in R

(Lakens, 2017). We used bounds of Cohen's d = 0.51 (Study 1) or d = 0.53 and d = 0.54 (Study 2), where power to detect an effect in the included group of participants is estimated to be 80%.

End-state learning accuracy in Study 1 averaged across the last 5 of 6 repetitions in the scanned intermixed learning session. End-state learning accuracy for Study 2 averaged across the last 2 of 3 repetitions in the final intermixed learning phase. For the purpose of correlations with working memory, initial learning repetitions 2-10 were averaged (as repetition 1 cannot reflect learning). In Study 1, the post-learning ratings were taken from the pre-scan ratings (following learning repetition 25).

Results

Across two studies, we measured learning and maintenance of conditioned stimulus-value associations over time. In the first in-lab session, participants learned stimulus-value associations for a set of "spaced-trained" stimuli. Over the course of the next two weeks, participants engaged in three further short training sessions online. Participants then returned to complete a second in-lab session, where they learned stimulus-value associations for a new set of "massed-trained" stimuli. All learning for the massed-trained stimuli occurred consecutively in the same session. By the end of training on the massed-trained stimuli, experience was equated between the spaced- and massed-trained stimuli. While the timing of trials was equivalent across the spaced-trained and massed-trained stimuli, the critical difference was that multiple days were inserted in-between training sessions for spaced-trained stimuli.

Study 1

Learning of value associations. Participants rapidly acquired the best "Yes" or "No" response for the reward- or non-reward-associated stimuli during learning. Within the first 3 repetitions of each stimulus, accuracy quickly increased to 77.0 % (95% Confidence Interval (CI) [73.7 80.2]) for spaced-trained stimuli and 79.0 % (CI [75.9 82.0]) for massed-trained stimuli (p-values < 0.001). By the end of the initial training section (repetition 10), performance increased to 82.9 % (CI [76.4 89.4]) for the spaced-trained stimuli and 93.2 % (CI [89.8 96.5]) for the massed-trained stimuli (**Figure 2a**). Performance was moderately higher by the end of initial learning for the massed-trained stimuli ($t_{(32)}$ = 3.13, CI [0.036 0.170]; p = 0.0037). However, as expected, after further experience we found that by the end of training participants no significant difference in performance across conditions (repetition 26-31; spaced-trained, 91.4 % CI [87.1 95.7]; massed-trained, 93.6 % CI [89.8 97.4]; $t_{(32)}$ = 1.34, CI [-0.012 0.055], p = 0.19; **Figure 2a**). However, this effect was statistically not equivalent to no effect, as indicated by an equivalence test using the TOST procedure (Lakens, 2017): the effect was not significantly within the bounds of a medium effect of interest (Cohen's *d* = ± 0.51, providing 80% power with 33 participants; $t_{(32)}$ = 1.59, p = 0.061), and thus we cannot reject the presence of a medium-size effect.

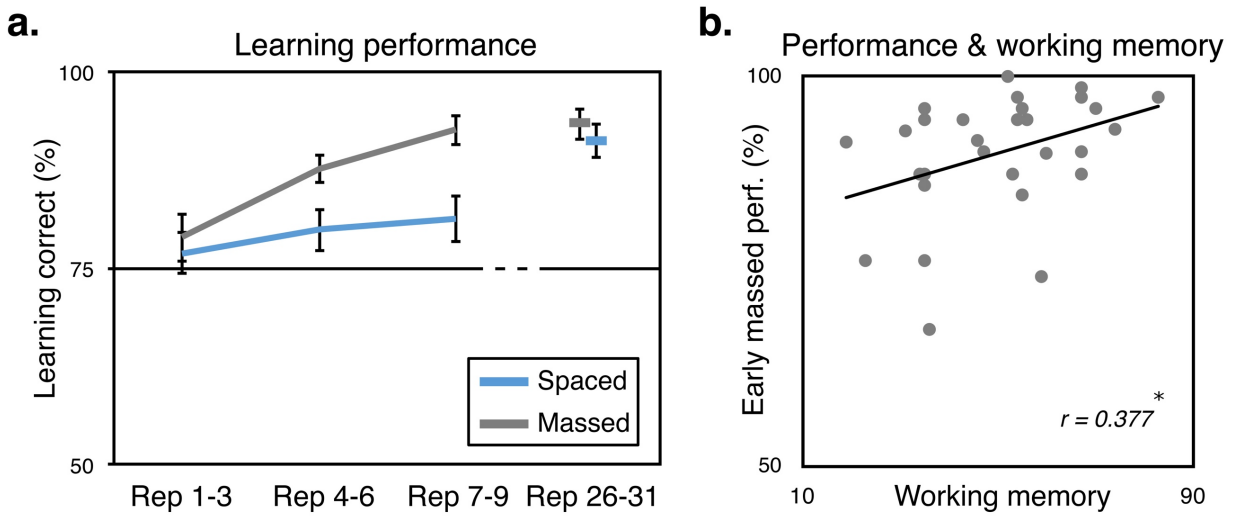

Figure 2. Study 1 learning results. a) Initial and endpoint learning performance for the spaced- and massed-trained stimuli. b) Positive trending correlation between early massed-trained stimulus learning phase performance and working memory capacity (O-SPAN). (∗ $p < 0.05$). Rep. = repetition. Error bars represent one standard error of the mean (s.e.m.).

After sufficient general experience in the task, we expected to find a positive relationship between learning performance for new stimuli and working memory. We thus estimated the correlation between learning during the initial acquisition of massed-trained stimulus-value associations during the second in-lab session with the operations span measure of working memory. We found that learning performance positively related to working memory capacity ($r = 0.377$, $p = 0.044$; **Figure 2b**). While significant at the $p < 0.05$ level, it is possible that our power to detect a stronger relationship with working memory was decreased by the near-ceiling level of performance in the reward task for many participants. Initial performance for spaced-trained stimuli did not correlate with working memory ($r = 0.031$, $p = 0.86$; TOST equivalence test, $p = 0.037$, providing 80% power in range $r \pm 0.34$). We did not predict a relationship between spaced condition performance and working memory because early during instructed tasks, working memory may be primarily taxed by maintaining task instructions (Cole et al., 2013). Similarly, while working memory capacity likely contributed to performance, initial task performance is also affected by numerous other noise-introducing factors

such as the acquisition of general task rules ("task set") and adaptation to the testing environment.

Long-term maintenance. Next, we turned to the critical question of whether spaced training over weeks led to differences in long-term memory for conditioned reward associations. Ratings were collected before the fMRI session and again at the end of the second in-lab session. High ratings indicate strong confidence in a reward association while low ratings indicate higher confidence in a neutral/loss association; ratings more toward the middle of the scale indicated less confidence (**Figure 1c**). After training but before fMRI scanning, when experience was matched across the spaced and massed conditions, we found that ratings across condition clearly discriminated between reward- and non-reward-associated stimuli (p-values < 0.001; **Figure 3a**, left). The difference between reward vs. non-reward ratings was larger in the massed than the spaced condition (spaced difference, 50.3 % CI [40.5 60.1]; massed difference, 62.0 % CI [55.5 68.6]; $t_{(30)}$ = 2.52; CI [2.2 21.2]; p = 0.017). Supporting the use of reward associations ratings as a measure of value in the long-term follow-up, we found that massed-trained stimulus ratings were strongly correlated with preferences in the separate choice test phase (mean r = 0.93 CI [0.90 0.96]; t-test on z-transformed correlation, $t_{(30)}$ = 11.06; CI [1.82 2.64]; p < 0.001).

To measure long-term maintenance of conditioning, after approximately 3 weeks, participants completed an online questionnaire on reward association strength using a 10-point scale. The instructions for ratings were the same as the in-lab ratings phase. Critically, we found that while the reward value discrimination was significant in both conditions (spaced difference, 4.46 CI [3.64 5.29]; $t_{(32)}$ = 11.08, p < 0.001; massed difference, 2.20 CI [1.51 2.90]; $t_{(32)}$ = 6.44, p < 0.001), reward value discrimination was significantly stronger in the spaced than in the massed condition ($t_{(32)}$ = 4.62; CI [1.26 3.26]; p < 0.001; **Figure 3b**). This effect was driven by the reward-associated stimuli (spaced vs. massed, $t_{(32)}$ = 4.91; CI [1.08 2.60]; p < 0.001; non-reward spaced vs. massed, $t_{(32)}$ = -1.33; CI [-1.07 0.23]; p = 0.19; TOST equivalence test, $t_{(32)}$ = 1.55, p = 0.066, n.s.).

Although the 3-week follow-up rating was collected on a 10-point scale and the post-learning ratings were collected on a graded scale, preventing direct numeric

comparison, the initial ratings and follow-up ratings can be analyzed with an across-time correlation. Such an analysis can test whether ratings in the massed case were simply scaled down (preserving an across-time correlation) or if actual forgetting introduced noise (disrupting an across-time correlation). We predicted that the value association memory for massed-trained stimuli actually decayed, leading to a higher correlation across time for spaced-trained stimuli. We indeed found that ratings were significantly more correlated across time in the spaced-trained condition (spaced r = 0.75 CI [0.64 0.87]; massed r= 0.45 CI [0.33 0.57]; t-test on z-transformed values, $t_{(30)}$ = 4.46; CI [0.51 1.38]; $p < 0.001$). We also collected the BDI as a measure of depressive symptoms in a subset of participants (n = 30); however, scores in this group were quite low (median = 3) and the distribution was strongly skewed toward zero, and thus we did not examine any relationship between BDI and learning. Overall, these results indicate that spaced-trained stimuli exhibited significantly stronger long-term memory for conditioned associations and more stable memory than massed-trained stimuli.

One limitation to these results is that in the current design, cues in the learning environment may bias performance in favor of the spaced-trained stimuli: online training for spaced stimuli was conducted outside the lab, likely on the participant's own computer, which was likely the same environment for the 3-week follow-up measure. While it seems unlikely that a testing environment effect would fully account for the large difference in long-term maintenance that we observed, we conducted a second study to replicate these results in a design where the testing conditions would if anything bias performance in favor of the massed-trained stimuli.

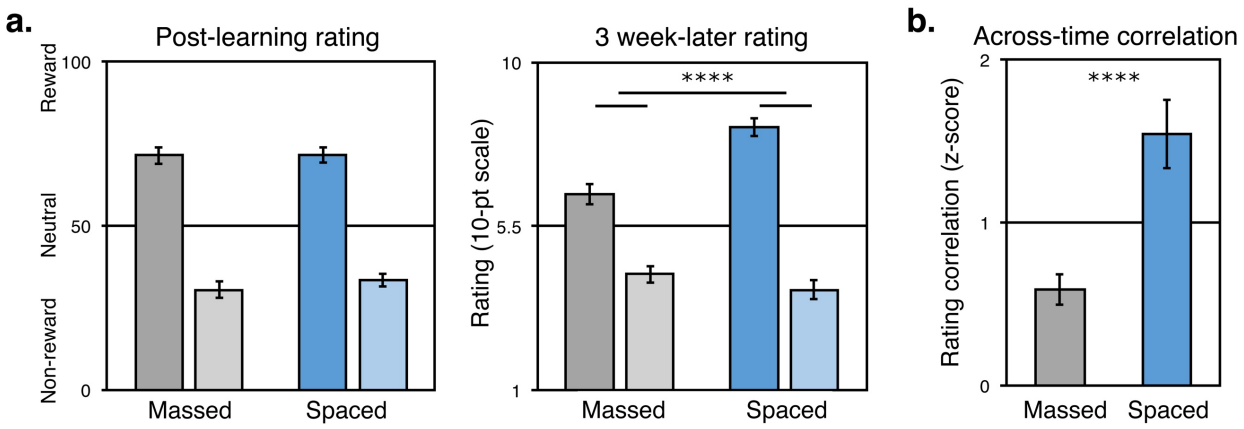

Figure 3. Study 1 post-learning value association strength and long-term maintenance of value associations. a) Post-learning reward association ratings for the massed- and spaced-trained stimuli (left); 3-week-later reward association ratings (right). Reward-associated stimuli in darker colors. b) Average of the correlation within-participant of massed-trained stimulus reward ratings and spaced-trained stimulus reward-ratings (z-transformed). Error bars, s.e.m.

Study 2

Learning of value associations. In Study 2, our aim was to replicate the findings of Study 1 and to extend them by conducting the 3-week follow-up session in the lab, allowing for a direct comparison with post-learning performance. During learning, within the first 3 trials, accuracy rapidly increased to 77.9 % (CI [74.8 81.0]) for spaced-trained stimuli and to 79.2 % CI [76.0 82.4] for massed-trained stimuli (p-values < 0.001). By the end of the initial training session, performance was at a level of 84.3 % (CI [79.3 89.3]) for the spaced-trained stimuli and 86.2 % (CI [81.2 91.1]) for the massed-trained stimuli (**Figure 4a**), which was matched across conditions (10^{th} repetition; $t_{(30)} = 0.59$, $p = 0.56$; TOST equivalence test within a range of Cohen's $d = \pm 0.53$, providing 80% power with 31 participants; $t_{(30)} = 2.37$, $p = 0.012$). By the end of training, after the online sessions for spaced-trained stimuli and the completion of the in-lab learning for massed-trained stimuli, we found that performance was equivalent across conditions (spaced-trained, 86.4 % CI [82.2 90.6]; massed-trained, 87.1 % CI [81.4 92.8]; $t_{(30)} = 0.248$, $p = 0.806$; TOST equivalence test, $t_{(30)} = 2.70$, $p = 0.006$; **Figure 4a**).

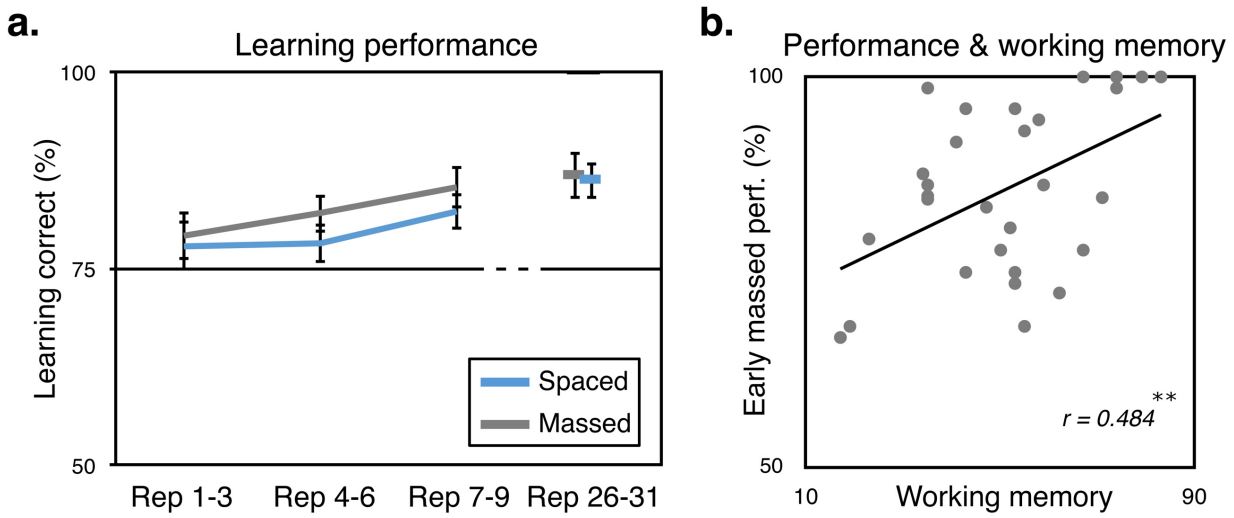

Figure 4. Study 2 learning results. a) Initial and endpoint learning performance for the spaced- and massed-trained stimuli. b) Positive correlation between early massed-trained stimulus learning performance and working memory capacity (OSPAN). (** $p < 0.01$). Error bars, s.e.m.

As in Study 1, after sufficient general experience in the reward association learning task, we expected to find a positive relationship between performance on the reward association learning task and working memory. Indeed, we found a significant correlation between massed-stimulus performance and working memory capacity ($r = 0.484$, $p = 0.0058$; **Figure 4b**). Initial learning performance was relatively lower in Study 2 than in Study 1, which may have helped reveal a numerically stronger correlation between massed-trained stimulus performance and working memory. Meanwhile, the relationship between working memory and initial performance for spaced-trained stimuli was weak ($r = 0.040$, $p = 0.83$; TOST equivalence test, $p = 0.043$, providing 80% power in range $r \pm 0.35$), as expected, given the other noise-introducing factors in initial learning performance discussed above.

Long-term maintenance. Next, we turned to the critical question of whether spaced training over weeks led to differences in long-term memory for conditioned reward associations. Ratings were collected at the end of the massed-stimulus training session but before fMRI scanning. With training experience matched, we found that ratings

across condition clearly discriminated between reward- and non-reward-associated stimuli (p-values < 0.001; **Figure 5a**, left). The difference between values tended to be higher in the massed than the spaced condition (massed difference, 52.5 % CI [46.7 58.3]; spaced difference, 47.1 % CI [40.9 53.2]; difference, $t_{(30)}$ = -1.86; CI [-0.5 11.4]; p = 0.073; **Figure 5a**). In the next phase, we found that incentive compatible choices overall were near ceiling (93.1% CI [89.9 96.4]; $t_{(30)}$ = 26.80) and did not differ between conditions (p > 0.26; TOST equivalence test, t(30) = 1.83, p = 0.04). Supporting the use of ratings in the long-term follow-up, we found that ratings positively correlated with choice preference across all stimuli (mean r = 0.87 CI [0.82 0.91]; t-test on z-transformed ratings, $t_{(30)}$ = 14.08; CI [1.33 1.78]; p < 0.001).

To measure long-term maintenance of conditioning, after approximately 3 weeks, participants returned for a third in-lab session. Rating discrimination between reward- and non-reward-associated stimuli was significant in both conditions (spaced difference, 39.1 % CI [32.4 45.8]; $t_{(29)}$ = 11.96, p < 0.001; massed difference, 16.7 % CI [9.4 24.1]; $t_{(29)}$ = 4.65, p < 0.001). Importantly, reward value discrimination was significantly stronger in the spaced than in the massed condition ($t_{(29)}$ = 4.98 CI [13.2 31.5], p < 0.001; **Figure 5a, right**). At follow-up, this stronger maintenance of learned value associations in the spaced condition was significant for both reward and non-reward stimuli (reward, $t_{(29)}$ = 3.43 CI [5.0 20.0], p = 0.0018; non-reward, $t_{(29)}$ = -4.11 CI [-14.7 -5.0], p < 0.001). The design of Study 2 allowed us to directly compare post-learning ratings and 3-week later ratings to calculate the degree of maintenance of conditioning. As expected, the difference in maintenance for reward associations was significantly greater for spaced- than massed-trained stimuli (spaced, 87.3 % CI [73.2 101.5]; massed, 30.0 % CI [16.2 43.9]; $t_{(29)}$ = 5.49 CI [36.0 78.6]; **Figure 5b**). Moreover, we found that ratings significantly decayed toward neutral for both reward- and non-reward-associated massed-trained stimuli (massed reward, $t_{(29)}$ = -6.09 CI [-21.7 -10.8], p < 0.001; non-reward, $t_{(29)}$ = 9.95 CI [15.3 23.3], p < 0.001). For spaced-trained stimuli, we found no decay for reward-associated stimuli but some decay for non-reward-associated stimuli (spaced reward, $t_{(29)}$ = -1.21 CI [-4.0 1.0], p = 0.23; TOST equivalence test, $t_{(29)}$ = 1.74, p = 0.045; non-reward, $t_{(29)}$ = 3.00 CI [2.1 11.4], p = 0.0055).

Finally, as in Study 1, we predicted that the value association memory for massed-trained stimuli was not decreased by scaling but actually decayed, which would lead to a lower across-time correlation in ratings. To test this, we correlated ratings in the second in-lab session with ratings in the third in-lab session separately for massed- and spaced-trained stimuli. We replicated the finding that ratings were significantly more correlated across time in the spaced-trained condition (spaced r = 0.82 CI [0.74 0.90]; massed r = 0.50 CI [0.37 0.63]; t-test on z-transformed values, $t_{(29)}$ = 5.22 CI [0.45 1.03], $p < 0.001$).

By collecting the long-term follow-up ratings in the same lab environment as the massed training sessions, our design would, if anything, be biased to find stronger maintenance for massed-trained stimuli because the training and testing environments overlap. However, we found similar differences in long-term conditioning across Study 1 and Study 2, suggesting that testing environment was not a significant factor in our measure of conditioning maintenance. The replication and extension of the findings of Study 1 provide strong evidence that spaced training leads to more robust maintenance of conditioned value associations at a delay, while performance in short-term learning is partly explained by working memory.

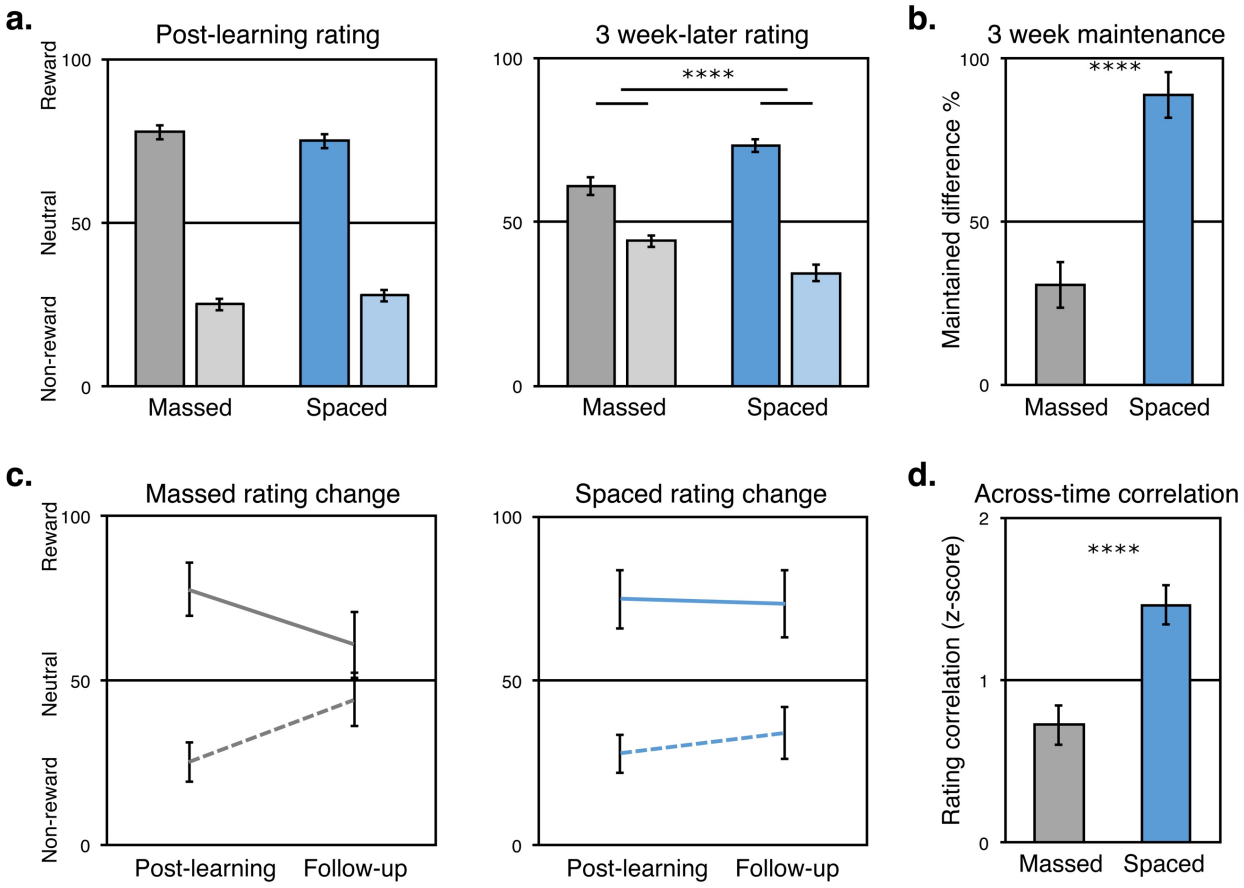

Figure 5. Study 2 post-learning reward association strength and maintenance of value associations. a) Reward association ratings for the massed- and spaced-trained stimuli after the second in-lab session (left), and after the 3-week-later in-lab final reward association rating session (right). b) Percent of initial reward association difference (reward minus non-reward associated rating) after the second in-lab session maintained across the 3-week delay to the third in-lab session, separately for massed- and spaced-trained stimuli. c) Post-learning and 3-week follow-up ratings re-plotted within condition for reward-associated (solid line) and non-reward-associated stimuli (dotted line). d) Average of the correlation within-participant of massed-trained stimulus reward ratings and spaced-trained stimulus reward-ratings (z-transformed). Error bars, s.e.m. (a, b, d), and within-participants s.e.m (c).

Discussion

When reward-based learning is distributed over time instead of massed in a single session, we found significant gains in maintenance of learned value associations. Controlling for the amount of training and post-training performance, across two experiments we found that stimuli trained across weeks exhibited significantly stronger maintenance of value associations in a surprise test 3 weeks later. Conversely, single-session massed training, as common employed in human experimental research, results in to weaker maintenance of value associations. This weaker maintenance may be related to greater reliance on short-term memory during massed learning, as we found that initial learning performance was significantly correlated with individual differences in working memory. Together, these results indicate that reward associations acquired from a single condensed session of learning will be less effective at guiding choices adaptively in the future. In order to understand the cognitive and neural mechanisms of habitual reward associations, including the potential influence of mood disorders and compulsive behaviors, it may be important to develop alternative paradigms that rely less on working memory and more on long-term reward association learning.

Research on learning and decision making has focused on two extreme timescales: short-term learning from reward feedback across minutes, for example, in “bandit” tasks (Daw et al., 2006), or choices based on well-learned values, for example, in snack food choices (Plassmann et al., 2007). There has been remarkably little research in humans that examines how well-learned value associations are actually established (Tricomi et al., 2009), even though our preferences are often shaped across days, months, or years of experience. Reward-based learning is known to depend on the striatum and its midbrain dopaminergic projections (Schultz et al., 1997; Rangel et al., 2008; Steinberg et al., 2013). It is possible that condensed single-session learning is primarily supported by the same neural mechanisms that support long-term and more habitual learning. However, recent findings strongly suggest that learning performance in tasks with condensed repetitions of stimuli is better captured by a reinforcement learning model augmented with a working memory component (Collins and Frank, 2012; Collins et al., 2014). Our results support the hypothesis that massed training in part relies on different neural systems than long-term training by demonstrating that

initial learning performance is also related to basic individual differences in working memory capacity, and further, that values learned in a massed session show significant decay over time.

What mechanisms support the improvement in long-term maintenance of values with spaced training? Previous research indicates several different, non-exclusive, mechanisms. At a cognitive level of analysis, a promising hypothesis is that spaced feedback learning is supported by additional neural mechanisms that, in normal situations, complement habitual value learning mechanisms in the striatum. One likely additional system supporting spaced learning is a memory system including the hippocampus. When learning events are spaced, an agent is forced to retrieve option values from longer-term memory (Bouton and Moody, 2004). Repeated retrieval in spaced conditions may itself support stronger learning: in another domain, repeated retrieval in an educational setting is known to increase test performance (Karpicke and Roediger, 2008).

At a circuit level, the hippocampus also presents a mechanism that may support beneficial effects of spacing. During navigation, activity in hippocampal cells represents the location of the animal in the environment, while during brief pauses or rest, rapid replay of these patterns of activity represent past and future paths to rewards (Johnson and Redish, 2007; Pfeiffer and Foster, 2013). Targeted online disruption of potential replay events significantly slows maze learning in rats (Jadhav et al., 2012). Further, striatal activity representing rewards occurs at expected timepoints in hippocampal replay events (Lansink et al., 2009; van der Meer and Redish, 2009). These findings suggest that coordinated replay in the hippocampus and striatum may support feedback learning.

Computationally, learning by offline replay fits well with the DYNA-Q model, where habitual values are trained by post-event replay or simulation of experience (Sutton, 1990; Johnson and Redish, 2005; Gershman et al., 2014). While such a mechanism has not been tested in simple reward-association learning in humans, episodic memory research has shown that the strength of post-event activity in the hippocampus and striatum is related to successful memory formation (Ben-Yakov and Dudai, 2011; Ben-Yakov et al., 2013). These findings suggest that spaced learning may benefit from active post-feedback processes that complement habitual value learning

mechanisms in the striatum. In typical rapid learning paradigms, such processes may be masked or inhibited.

At the cellular level, it is possible that processes supporting learning may operate more effectively with more time between learning events (Reynolds et al., 2001). Long-term plasticity (LTP), which is believed to support learning in neural systems, involves an essential late phase, including nuclear transcription and even the recruitment of new synapses, which can take from minutes to hours (Woo et al., 2003; Kramar et al., 2012; Aziz et al., 2014). Post-learning consolidation that includes sleep may be particularly advantageous, based on what is known about the effects of sleep in other domains (Walker and Stickgold, 2006). By spacing training over long time periods, the cellular processes underlying LTP processes can be re-engaged multiple times. Indeed, early studies of LTP demonstrate positive effects of spacing (Carew et al., 1972). Spaced training may establish a larger number of potentiated synapses and/or a more efficient storage of the memory, which may be more resistant to decay over time.

During massed learning, other cognitive mechanisms may be engaged during that mask or impair normal reward-based learning mechanisms. One likely mechanism is short-term or working memory, which can maintain a representation of the recent past. During a reward-based learning task, when events are close together in time, working memory can in principle support the maintenance of choice- and feedback-related information to guide the next choice. Such short-term representations of recent stimuli can lead to different learning processes, as these maintained representations may block the retrieval of associations from longer-term memory (Bouton and Moody, 2004). In humans, recent work has demonstrated that working memory likely plays an important role in feedback learning, over and above the role of traditional mechanisms of habitual value learning (Collins and Frank, 2012; Collins et al., 2014). In these experiments, the authors examined simple stimulus-reward association learning when participants needed to learn the correct (rewarded) responses for different numbers of stimuli in a set. This manipulation of set size effectively increases the spacing between repetitions of a given stimulus. Contrary to predictions from simple reinforcement learning models, participants exhibited significantly slower learning as set size increased. To explain the negative effect of set size on learning, the authors developed an augmented reinforcement learning model that included a working memory

component. Using this model, in a study of patients with Schizophrenia, the authors found that deficits in performance were primarily due to differences in the working memory parameter but not reward learning parameters (Collins et al., 2014).

The mechanisms described above may also lead to differences in how well-learned vs. recently-learned reward associations represented in the brain. Decades of animal research have shown that different regions of the striatum are important for different types of reward associations, with dorsomedial striatal regions critical for flexible (and newly-acquired) goal-directed learning and the dorsolateral striatum critical for inflexible habit learning (Balleine and Dickinson, 1998; Yin and Knowlton, 2006). In conjunction with the development of representations in the dorsolateral striatum, habits frequently come to dominate behavior over time. Recently, in non-human primates it has been reported that months-long versus single-session learning of reward associations leads to the establishment of reward associations in distinct striatal and midbrain regions (Kim and Hikosaka, 2013; Kim et al., 2015). In the striatum, anterior caudate neurons reflected recently-learned and switching values, while neurons in the caudate tail reflected stable value associations learned over weeks and months (Kim and Hikosaka, 2013). Critically, even after stimulus-reward associations were extinguished, the authors found that a novel population of dopamine neurons persisted in responding to the long-extinguished stimuli, an effect that correlated with residual attentional bias. This work suggests that attentional orientation to reward-associated cues may be considered a “habit” that is resistant to extinction (Kim et al., 2015; Anderson, 2016).

In humans, an initial demonstration that habit-like responding can be studied in the brain was reported by Tricomi and colleagues (Tricomi et al., 2009). Here, participants engaged in reward association learning across multiple days of scanning. The authors report that training led to participant’s being less sensitive to devaluation, a hallmark of habitual behavior (Dickinson and Balleine, 2002), and to an increase in activity in the putamen, a region believed to be homologous to the rodent dorsolateral striatum, increased over time (Tricomi et al., 2009). Multi-day training was also employed by Wunderlich and colleagues (Wunderlich et al., 2012), where stimulus-reward associations learned over days led to value-correlated responses in the putamen. These results suggest that in humans, similar to research in animals, the

posterior striatum may be important for encoding well-learned values. However, these experiments did not compare multi-day training to matched massed training, and posterior striatal activity is not specific, as reward-related responses in the posterior striatum have also been reported for single-session fMRI experiments (O'Doherty et al., 2003; Dickerson et al., 2011; Wimmer et al., 2014). Thus, the neural signatures specific to well-learned vs. transient value associations in the human brain remain unknown.

Our results have implications for understanding reward-based learning in the healthy brain and for translating this research to patient populations (Huys et al., 2016). Many reward-based learning studies examine behavior and brain activity in rapidly-paced probabilistic choice tasks (e.g. Pessiglione et al., 2006) or drifting “bandit” paradigms (e.g. Daw et al., 2006; Wimmer et al., 2012; Doll et al., 2016). Learning- and choice-related parameters estimated from computational models can then be related to differences in neural activity and individual differences in personality or group membership. However, the interpretation of parameters derived from massed training paradigms is difficult for several reasons. First, the optimality of faster or slower reward learning is dependent on the reward statistics of the environment (Daw, 2011), complicating generalizations based on a single task. Second, as we demonstrate, initial learning in a massed training session is significantly related to individual differences in working memory capacity, a finding supported by reinforcement learning models that include a working memory component (Collins and Frank, 2012). Third, we found that spaced training, which matches more closely the training paradigms used to study habit learning in animal research, results in significantly better maintenance of conditioned value associations. Thus, characteristics of reward-based learning in massed paradigms are not likely to be pure indications of individual differences in reward learning outside the lab. For translational research in learning and in cognitive neuroscience in general, it will be important to develop and utilize more ecologically valid experimental designs (Moutoussis et al., 2016).

By demonstrating that we can establish and measure long-term reward associations that are resistant to decay, our experimental design provides a starting point for studying behavioral change. We have provided an initial demonstration that long-term conditioned associations can be studied in a controlled manner. In future translational research, it will be important to examine how well-learned value

associations and habits of behavior be adjusted or unlearned. Well-learned behaviors are notoriously difficult to modify. As noted above, the maintenance of long-term reward associations may also be reflected in a qualitatively different neural response that persists after devaluation, distinct from the well-known reward prediction error response in dopamine neurons (Schultz et al., 1997; Kim et al., 2015).

In summary, across two studies we found that spacing of reward-based learning across weeks versus minutes results in significantly greater maintenance of conditioned value associations. We also found that learning performance in the massed condition was related to individual differences in working memory. Our experiments represent the first demonstration of spacing effects on reward-based learning in humans. While it is widely assumed that a habitual value learning system in the striatum supports feedback learning in recent behavioral and fMRI studies in humans, the beneficial effects of a more ecologically relevant spaced training condition suggest that current massed paradigms may be eliciting a mixture of different learning systems. Spaced reward-based learning and long-term maintenance of conditioning may thus help provide cleaner measures of feedback-based learning in the striatal dopamine system. This possibility has implications for the interpretation and future direction of reward-based learning research, as feedback learning paradigms are becoming widely used in studies of mood and psychiatric disorders as well as addiction (Herbener, 2009; Maia and Frank, 2011; Montague et al., 2012; Whitton et al., 2015). Our study of healthy adults points to potential promising avenues to explore in order to more fully understand the cognitive and neural learning mechanisms of feedback-based learning.

References

Anderson BA (2016) The attention habit: how reward learning shapes attentional selection. Ann N Y Acad Sci 1369:24-39.

Aziz W, Wang W, Kesaf S, Mohamed AA, Fukazawa Y, Shigemoto R (2014) Distinct kinetics of synaptic structural plasticity, memory formation, and memory decay in massed and spaced learning. Proc Natl Acad Sci U S A 111:E194-202.

Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology 37:407-419.

Barto AG (1995) Adaptive critics and the basal ganglia. In: Models of Information Processing in the Basal Ganglia (Davis JL, Houk JC, Beiser DG, eds), pp 215-232. Cambridge: MIT Press.

Ben-Yakov A, Dudai Y (2011) Constructing realistic engrams: poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. J Neurosci 31:9032-9042.

Ben-Yakov A, Eshel N, Dudai Y (2013) Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. J Exp Psychol Gen 142:1255-1263.

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

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433-436.

Carew TJ, Pinsker HM, Kandel ER (1972) Long-term habituation of a defensive withdrawal reflex in aplysia. Science 175:451-454.

Carpenter SK, Cepeda NJ, Rohrer D, Kang SHK, Pashler H (2012) Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. Educ Psychol Rev 24:369-378.

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.

Cole MW, Laurent P, Stocco A (2013) Rapid instructed task learning: a new window into the human brain's unique capacity for flexible cognitive control. Cogn Affect Behav Neurosci 13:1-22.

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.

Daw ND (2011) Trial-by-trial data analysis using computational models. In: Attention & Performance XXII (Delgado MR, Phelps EA, Robbins TW, eds), pp 3-38. Oxford, UK: Oxford University Press.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876-879.

de Leeuw JR (2015) jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. Behav Res Methods 47:1-12.

Dickerson KC, Li J, Delgado MR (2011) Parallel contributions of distinct human memory systems during probabilistic learning. Neuroimage 55:266-276.

Dickinson A, Balleine B, eds (2002) The role of learning in motivation, 3 Edition. New York: Wiley.

Doll BB, Bath KG, Daw ND, Frank MJ (2016) Variability in Dopamine Genes Dissociates Model-Based and Model-Free Reinforcement Learning. J Neurosci 36:1211-1222.

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

Eichenbaum H, Cohen NJ (2001) From Conditioning to Conscious Recollection: Memory Systems of the Brain. New York: Oxford University Press.

Ellenbogen JM, Hu PT, Payne JD, Titone D, Walker MP (2007) Human relational memory requires time and sleep. Proc Natl Acad Sci U S A 104:7723-7728.

Garrison J, Erdeniz B, Done J (2013) Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. Neuroscience and biobehavioral reviews 37:1297-1310.

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.

Herbener ES (2009) Impairment in long-term retention of preference conditioning in schizophrenia. Biol Psychiatry 65:1086-1090.

Houk JC, Adams JL, Barto AG (1995) A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: Models of information processing in the basal ganglia (Houk JC, Davis JL, Beiser DG, eds), pp 249-270. Cambridge, MA: MIT Press.

Humphreys LG (1940) Distributed practice in the development of the conditioned eyelid response. J Gen Psychol 22:379-385.

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

Jadhav SP, Kemere C, German PW, Frank LM (2012) Awake hippocampal sharp-wave ripples support spatial memory. Science 336:1454-1458.

Johnson A, Redish AD (2005) Hippocampal replay contributes to within session learning in a temporal difference reinforcement learning model. Neural Netw 18:1163-1171.

Johnson A, Redish AD (2007) Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. J Neurosci 27:12176-12189.

Karpicke JD, Roediger HL, 3rd (2008) The critical importance of retrieval for learning. Science 319:966-968.

Kim HF, Hikosaka O (2013) Distinct basal ganglia circuits controlling behaviors guided by flexible and stable values. Neuron 79:1001-1010.

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

Kim JJ, Thompson RF (1997) Cerebellar circuits and synaptic mechanisms involved in classical eyeblink conditioning. Trends Neurosci 20:177-181.

Knowlton BJ, Mangels JA, Squire LR (1996) A neostriatal habit learning system in humans. Science 273:1399-1402.

Kramar EA, Babayan AH, Gavin CF, Cox CD, Jafari M, Gall CM, Rumbaugh G, Lynch G (2012) Synaptic evidence for the efficacy of spaced learning. Proc Natl Acad Sci U S A 109:5121-5126.

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

Lansink CS, Goltstein PM, Lankelma JV, McNaughton BL, Pennartz CM (2009) Hippocampus leads ventral striatum in replay of place-reward information. PLoS Biol 7:e1000173.

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

Litman L, Davachi L (2008) Distributed learning enhances relational memory consolidation. Learn Mem 15:711-716.

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

Menzel R, Manz G, Menzel R, Greggers U (2001) Massed and spaced learning in honeybees: The role of CS, US, the intertrial interval, and the test interval. Learn Memory 8:198-208.

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.

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003) Temporal difference models and reward-related learning in the human brain. Neuron 38:329-337.

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.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442:1042-1045.

Pfeiffer BE, Foster DJ (2013) Hippocampal place-cell sequences depict future paths to remembered goals. Nature 497:74-79.

Plassmann H, O'Doherty J, Rangel A (2007) Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. J Neurosci 27:9984-9988.

Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci 9:545-556.

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

Roesch MR, Calu DJ, Schoenbaum G (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci 10:1615-1624.

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.

Schultz W (2011) Potential vulnerabilities of neuronal reward, risk, and decision mechanisms to addictive drugs. Neuron 69:603-617.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593-1599.

Sochat VV, Eisenberg IW, Enkavi AZ, Li J, Bissett PG, Poldrack RA (2016) The Experiment Factory: Standardizing Behavioral Experiments. Frontiers in psychology 7:610.

Spence KW, Norris EB (1950) Eyelid conditioning as a function of the inter trial interval. Journal of Experimental Psychology 40:716-720.

Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH (2013) A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci 16:966-973.

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.

Sutton RS, Barto AG (1998) Reinforcement Learning: an Introduction. Cambridge: MIT Press.

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.

Tricomi E, Balleine BW, O'Doherty JP (2009) A specific role for posterior dorsolateral striatum in human habit learning. Eur J Neurosci 29:2225-2232.

van der Meer MA, Redish AD (2009) Covert expectation-of-reward in rat ventral striatum at decision points. Front Integr Neurosci 3:1.

Walker MP, Stickgold R (2006) Sleep, memory, and plasticity. Annu Rev Psychol 57:139-166.

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, Shohamy D (2012) Preference by association: how memory mechanisms in the hippocampus bias decisions. Science 338:270-273.

Wimmer GE, Daw ND, Shohamy D (2012) Generalization of value in reinforcement learning by humans. Eur J Neurosci 35:1092-1104.

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.

Woo NH, Duffy SN, Abel T, Nguyen PV (2003) Temporal spacing of synaptic stimulation critically modulates the dependence of LTP on cyclic AMP-dependent protein kinase. Hippocampus 13:293-300.

Wunderlich K, Dayan P, Dolan RJ (2012) Mapping value based planning and extensively trained choice in the human brain. Nat Neurosci 15:786-791.

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