Reward Prediction Errors Shape Memory during Reinforcement Learning

Nina Rouhani

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
PSYCHOLOGY

Advisers: Yael Niv (primary) & Ken Norman (secondary)

June 2020

Abstract

In this dissertation, I characterize the role or reward prediction errors (RPEs) in shaping episodic memory across three series of behavioral experiments and computational modeling of learning and memory behavior. In Chapter 1, I show that large unsigned RPEs increase learning for those outcomes (i.e., learning rate) as well as memory for those outcome events. However, I do not find these effects to be correlated, suggesting distinct underlying mechanisms. In Chapter 2, I further test whether depressive symptoms modulate unsigned-RPE effects on learning and memory. I do not find depressive symptoms to lead to overall differences in learning and memory. Instead, I find that symptom group predicts opposite biases in the unsigned-RPE modulation of memory: in depressive participants, unsigned RPEs increased memory more for negative- versus positive-RPE events, whereas in non-depressive participants, unsigned RPEs increased memory more for positive- versus negative-RPE events. In Chapter 3, I dissociate the effects of RPEs experienced at reward cue from those at outcome on learning and memory for those events. I show, in line with classic associative models of attention, that signed RPEs at reward cue and unsigned RPEs at reward outcome modulate a dynamic learning rate in reinforcement learning models fit to behavior. When characterizing RPE effects on memory, I replicate previous results and find that unsigned RPEs at outcome enhance memory throughout learning, especially for outcome events. In addition to this, memory for cue events increases as a function of learning wherein a signed RPE at cue boosts memory for events associated with more valued reward categories. Finally, in Chapter 4, I investigate the computational mechanism supporting better memory for large unsigned-RPE events by testing whether they create event boundaries in memory. Large-RPE events are more strongly encoded and show intact associative links with their predecessors; nevertheless, they consistently disrupt the integration of events that occur across them, thereby creating event boundaries in memory. I capture these effects in a computational model of memory modified to incorporate RPEs into the encoding process. To conclude, I link my findings to interactions between reinforcement learning and memory systems, offering targets for future neuroscientific research.

Contents

This dissertation is dedicated to my grandmother, Soussan Shams, who encourages and reminds me every Sunday afternoon that she will not leave this world until I win a Nobel Prize, unknowingly providing the best excuse to never near one.

Acknowledgments

I would first and foremost like to thank my advisers Yael Niv and Ken Norman whose thoughtful and relentless dedication to their students and their science demonstrates a standard of mentorship I previously thought impossible, and frankly, still think supernatural. Yael, your untiring support of my work in every one of its stages (from showing me how a for loop works to editing this manuscript) has shaped me as a scientist, bottom-up and top-down. Your open-mindedness and enthusiasm for research gave me the excitement to pursue any question, while your unfaltering empirical approach and exacting attention to detail funneled that excitement into producing building lines of research. Ken, although technically my second adviser, you made yourself as available as a first. The unhesitating confidence you gave me (and all your students) to pursue, at first glance, scarier tasks, pushed me to develop skills I previously lacked some courage for. Challenges were further alleviated by your uncanny ability to make abstract, complex relationships, tangible (e.g., semantic bananas). Yael and Ken, your sincere, humble, and always constructive approach to supporting your students has made rocky times feel pebbly instead. I am immensely grateful for you both, as scientists and as human beings.

I also extend an immeasurable thank you to Daphna Shohamy, my first adviser in this field, without whom I would most certainly not have pursued this research. Your trailblazing approach to bridging work in reinforcement learning and memory has, quite obviously, inspired every chapter in this dissertation.

At Princeton, I could not imagine life without Yeon Soon Shin and Angela Radulescu. Yeon Soon, as my labmate and roommate, I have selfishly benefited from knowing where you are at most times. Your help and words and spirit have been the brighter (but not fluorescent) lights of my living spaces. Angela, we have journeyed through so many event boundaries, I can barely mental time-travel to a time I didn't know you, and for that I am thankful. I am also grateful to have collaborated in work (and in play) with Sarah DuBrow, Aaron Bornstein and James Antony. You've each made science so fun that not even drinks or deafening karaoke could get in the way of talking shop.

I am also enormously indebted to my family:

My parents, Shahrokh and Firouzeh, who unconditionally supported me through all of my career turns and taught me the emotional resilience that comes with living a life uprooted. I would quite literally not be a person, or more specifically, where I am today without my mom's unwavering strength and curiosity for the world, or without my dad's blistering intelligence, sense of humor, and incredible encouragement to pursue what brings me joy. I am obscenely lucky to have been raised by two champions for my safety and happiness, spotlit during this bizarre global pandemic.

My sister, Shiva, for whom I have no words to express the depth of my gratitude, but from the bottom of my being, you are my favorite piece of life.

My chosen family (in chronological order of when I chose them, not necessarily when I met them): Tiffany Peon, Julia Pierpont, Brian McElhaney, and Chris McIntyre, you remind me of the best parts of being human (it's talking to you).

And last but not least, I would very sincerely like to thank television, and all of its streaming-service vehicles, for keeping me going during long nights spent debugging code. In particular, I am beholden to three genres of TV shows: self-aware vampire dramedies targeted to preteens (e.g., 'Buffy the vampire slayer'), epic dramas targeted to disturbed adults (e.g., 'Six feet under', 'Succession'), and irreverent yet profound comedies targeted to me (e.g., 'Fleabag'). These shows have undoubtedly perpetuated my interest in reward prediction errors and subsequent memory for those events.

0

Introduction

Memories help guide future behavior, but which experiences from the past are prioritized in memory? In reinforcement learning, the value of a rewarding source is computed as a weighted average over all experienced outcomes, suggesting we integrate across multiple memories when making a decision. In contrast, episodic memories represent single events that allow for rapid, one-shot learning of relations between stimuli and outcomes. These two systems have been thought to engage separate underlying mechanisms, with the dopaminergic midbrain and striatum supporting reinforcement learning (for a review, see Schultz, 2015), and the hippocampus supporting episodic memory (for a review, see Moscovitch, Cabeza, Winocur, & Nadel, 2016). However, a large and growing literature has blurred this division, showing that learning and memory systems overlap and interact while predicting decision-making in a variety of behavioral and neural paradigms (Bornstein & Norman, 2017; Gershman & Daw, 2017; Shohamy & Daw, 2015; Wimmer, Daw, & Shohamy, 2012). For example, both the expected value of an option and a distinct memory of a previously experienced outcome influence a decision (Biele, Erev, & Ert, 2009; Duncan & Shohamy, 2016). Moreover, recent sampling models of decision-making posit that previous outcomes are drawn from memory to form the internal value of an option (Lieder, Griffiths, & Hsu, 2018; Shadlen & Shohamy, 2016; Bakkour et al., 2019; Bornstein & Norman, 2017; Bornstein, Khaw, Shohamy, & Daw, 2017). It is nevertheless unclear which experiences may be more or less available in memory. Here, I investigate how reward prediction errors – rapid and transient reinforcement signals that track the difference between actual and expected outcomes – shape our memory for rewarding events.

Reward prediction errors play a well-established role in updating stored information about the values of different choices, and are known to modulate dopamine release. When a reward is better than expected, there is an increase in the firing of dopamine neurons, and conversely, when the reward is worse than expected, there is a dip in dopaminergic firing below the baseline firing rate (Barto, 1995; Montague, Dayan, & Sejnowski, 1996). Dopamine, in turn, modulates plasticity in the hippocampus, a key structure for episodic memory (Lisman & Grace, 2005; Shohamy & Adcock, 2010). This dopaminergic link therefore provides a potential neurobiological mechanism for reward prediction errors to affect episodic memory. However, there are several ways by which reward prediction errors could potentially influence episodic memory.

First, if memory formation is affected by this *signed* reward prediction error, then we would expect an asymmetric effect on memory, such that a positive prediction error (leading to an increase in dopaminergic firing) would improve memory whereas a negative prediction error (leading to a decrease in dopaminergic firing) would worsen it (Jang, Nassar, Dillon, & Frank, 2019; Davidow, Foerde, Galvan, & Shohamy, 2016).

A second possibility is that the magnitude of the prediction error could influence episodic memory regardless of the sign of the error, enhancing memory for events that are either much better or much worse than expected. Such effects of *unsigned* prediction errors are thought to be mediated by the locus-coeruleus-norepinephrine (LC-NE) system, which demonstrates a transient response to unexpected changes in stimulus-reinforcement contingencies in both reward and fear learning (that is, regardless of sign; for a review, see Sara, 2009). Importantly, this LC-NE signal is thought to increase learning rate, i.e. the extent to which a learner updates their values, following large unsigned prediction errors (Behrens, Woolrich, Walton, & Rushworth, 2007; McGuire, Nassar, Gold, & Kable, 2014; Nassar et al., 2012; Pearce & Hall, 1980). Importantly, recent evidence also indicates that the locus coeruleus co-releases dopamine with norepinephrine, giving rise to dopamine-dependent plasticity in the hippocampus (Kempadoo, Mosharov, Choi, Sulzer, & Kandel, 2016; Takeuchi et al., 2016; Wagatsuma et al., 2017). This work highlights a new source of dopamine other than the VTA and substantia nigra pars compacta (where the signed reward prediction error originates), which leads to new hypotheses of how events that modulate LC activity, such as unsigned reward prediction errors, might boost hippocampal memories (Duszkiewicz, Mcnamara, Takeuchi, & Genzel, 2018; Hauser, Eldar, Purg, Moutoussis, & Dolan, 2019; Clewett & Murty, 2019; Clewett, Huang, Velasco, Lee, & Mather, 2018).

In **Chapter 1**, I first answer the question of whether and how reward prediction errors (henceforth RPEs) modulate learning and memory (Rouhani, Norman, & Niv, 2018). To do this, in three experiments, we investigated whether learning in a context characterized by high-outcome variance ('high-risk') with frequent, large RPEs, gives rise to higher fidelity memory traces than learning in a 'low-risk context'. In each experiment, we showed that large unsigned RPEs increased learning rate for outcomes and enhanced memory for those events. Nevertheless, we did not find a relationship between learning rate for reward and memory for items. In fact, on average, the high- versus low-risk context led to lower learning rates but better memory; that is to say, learning rates scaled to the outcome variance whereas memory did not. These results show that unsigned RPEs boost both incremental reward learning and episodic memory, but the two effects are likely mediated by distinct neural systems.

In **Chapter 2**, I re-analyze data from Chapter 1 to determine whether self-reported symptoms of depression modulate the effects of unsigned RPEs on learning and memory (Rouhani & Niv, 2019). We did not find overall differences in reward learning between subjects with depressive symptoms versus those without, but, within subjects that did report depressive symptoms, depression severity predicted worse learning. There were also no overall differences in memory performance across the groups; however, we found distinct biases for the unsigned-RPE enhancement of memory. In participants with depressive symptoms, large negative RPEs boosted memory more so than did large positive RPEs, and vice versa for non-depressive participants who showed more an effect of large positive (versus negative) RPEs in improving memory. In other words, individuals with depressive symptoms were more likely to remember surprising negative events, whereas those without depressive symptoms were more likely to remember surprising positive events.

In **Chapter 3**, I dissociate the effects of signed and unsigned RPEs experienced at reward cue versus outcome on memory for events presented at reward cue and outcome. In line with classic associative models of attention (Pearce & Mackintosh, 2010), we studied how learning dynamics influence memory for cue and outcome events. Here, we expected that an unsigned RPE signal at reward outcome would boost memory throughout learning for surprising events (as in Chapter 1, and Pearce & Hall's (1980) model of attention), whereas a signed RPE signal at the reward-predictive cue would increase memory once the values of different categories have been learned (as in Mackintosh's (1975) model of attention).

Across two experiments, participants learned the values of categories and, critically, saw two trial-unique images on every trial, one serving as the cue event and the other alongside the reward as the outcome event. We tested reinforcement learning models that included attentional components – an unsigned RPE at outcome (Pearce-Hall) and a signed RPE at cue (Mackintosh) – as modulators of trial-by-trial learning rates, and found these models to fit learning behavior better than models without those components. When characterizing RPE effects on memory, we replicated previous results (Chapter 1) and found that unsigned RPEs at outcome enhanced memory throughout learning, especially for events experienced at reward outcome. In addition to this, we found that memory for cue events increased as a function of learning wherein a signed RPE at cue boosted memory for events associated with more valued reward categories. Last, in a final choice test, we found that participants preferred both cue and outcome events linked to higher rewards and more valued reward categories. Interestingly, however, when choosing between a cue and an outcome event that belonged to the same trial (same reward and value; there should be no preference for either event), the signed RPE at outcome, which did not predict memory, increased preference for the outcome event, pointing to a hedonic influence of the signed RPE in modulating choice. In summary, we show here that signed and unsigned RPEs separately and dynamically influence learning and memory for rewarding events.

In **Chapter 4**, I investigate the computational mechanism supporting better memory for events accompanied by large unsigned RPEs (Rouhani, Norman, Niv, & Bornstein, 2020). Specifically, we tested whether this signal simply *strengthens* the encoding of events or whether it acts as an event boundary in memory, thereby improving memory by *separating* events that came after it from those that came before it in memory, thus reducing interference from those past events. Across four experiments, participants learned to predict rewards in distinct contexts. Within each context, high-magnitude RPEs ('high RPEs') indicated a shift in the underlying distribution of rewards. We used recognition priming and temporal-order memory to assess whether high RPEs create event boundaries by interrupting the temporal integration of events in memory.

When testing sequential items that did or did not contain a high-RPE event, we found recognition priming for both pair types; however, this effect was diminished when testing pairs *across* a high RPE. We also found and replicated intact sequence memory for sequential pairs that included a high-RPE event, but again found worse sequence memory for events across a high-RPE event, mirroring our recognition priming results. Moreover, greater distance between events at encoding led to better sequence memory for events that spanned a low-RPE event, but not a high-RPE event, suggesting separate mechanisms for the temporal ordering of events within versus across a latent reward context. Altogether, these findings demonstrate that high-RPE events are both more strongly encoded, show intact links with their predecessor, and yet act as event boundaries that disrupt the integration of sequential events in memory. We captured these effects in a variant of the Context Maintenance and Retrieval model (Polyn, Norman, & Kahana, 2009), modified to incorporate RPEs into the encoding process.

Finally, in the **Conclusion**, I bridge across the results of each study by offering putative neural mechanisms supporting our findings, and discuss some open and new questions.

1

Dissociable Effects of Unsigned Reward Prediction Errors on Learning and Memory[*]

We first tested whether signed or unsigned prediction errors influence learning rate and episodic memory, and whether these two effects are correlated. Correlated effects on learning of values and memory for events would suggest a common mechanism underlying both effects, whereas two uncorrelated effects are consistent with separate underlying mechanisms.

We also measured the effect of risk context (i.e., whether unsigned prediction errors were large or small, on average, in a particular environment) on episodic memory. Previous work on the effects of risk context show that dopamine signals scale to the reward variance of the learning environment (Tobler, Fiorillo, & Schultz, 2005), allowing for greater sensitivity to prediction errors in lower variance contexts. Moreover, behavioral learning rate and BOLD responses in the dopaminergic midbrain and striatum reflect this adaptation, with higher learning rates and increased striatal response to prediction errors when the reward variance is lower (Diederen, Spencer, Vestergaard, Fletcher, & Schultz, 2016). We therefore expected higher learning rates in a low-risk context, but it was unclear whether this effect would interact with episodic memory. If anything, for memory we expected opposite effects, such that a high-risk context would induce better episodic memory, as salient feedback (like experiencing high magnitude prediction errors) is thought to increase autonomic arousal and encoding of those events (Clewett, Schoeke, & Mather, 2014). The mnemonic effects of higher magnitude prediction errors may also "spill over" to surrounding items, boosting memory for those items as well, again predicting better memory for events experienced in the high-risk context (Duncan, Sadanand, & Davachi, 2012; Mather,

---

[*] The content of this chapter was published in Rouhani, Norman & Niv, 2018.

Clewett, Sakaki, & Harley, 2015).

To investigate the effect of reward prediction errors and risk context on the structure of memory, we asked participants to learn by trial and error which of two types of images, indoor or outdoor scenes, leads to larger rewards. Trial-unique indoor and outdoor images were presented in two different contexts or 'rooms,' with each room associated with a different degree of outcome variance. The average values of the scene categories in the two rooms were matched. Participants were instructed to learn the average (expected) value of each type of image (indoor or outdoor scenes), given the variable individual outcomes experienced for each scene, as is typically done in reinforcement learning tasks (e.g., O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003).

Specifically, we asked participants to explicitly estimate, on each trial, the average value of the category of the current scene. The deviation between this estimate and the outcome on that trial defined the trial-specific subjective prediction error. These prediction errors were then used to calculate trial-by-trial learning rates for the average values of the categories, as well as to predict future memory for the specific scenes presented on each trial. At a later stage, memory for the individual scenes was assessed through recognition memory ('item' memory), identification of the room the item belonged to ('source' or context memory; Exp. 2-3), and the ordering of a pair of items ('sequence' memory). Given that both category-value learning and individual scene memory were hypothesized to depend on the same prediction errors, we also characterized the relationship between learning about the average rewards in the task and episodic memory for the individual rewarding events.

1.1 Experiment 1

In Experiment 1, we assessed whether reward prediction errors interact with episodic memory for rewarding episodes. Participants learned the average reward values of images from two categories (indoor or outdoor scenes) in two learning contexts ('rooms'). The two learning contexts had the same mean reward, but different degrees of reward variance ('risk') such that the rewards associated with scenes in the 'high-risk room' gave rise to higher absolute prediction errors than in the 'low-risk room'. We then assessed participants' recognition for the different scenes in a surprise memory test, to test how prediction errors due to the reward associated with each episode affected memory for that scene.

Method

Participants

Two hundred participants initiated an online task using Amazon Mechanical Turk (MTurk), and 174 completed the task. We obtained informed consent online, and participants had to correctly answer questions checking for their understanding of the instructions before proceeding; procedures were approved by Princeton University's Institutional Review Board. Participants were excluded if they (1) had a memory score (A': Sensitivity index in signal detection, Pollack & Norman, 1964) of less than 0.5 based on their hit rate and false alarm rate for item recognition memory, or (2) missed more than three trials. These criteria led to the exclusion of ten participants, leading to a final sample of 164 participants. Although we do not have demographic information for the mTurk workers who completed these experiments, an online demographic tracker reports that during the time we collected data, the samples were approximately 55% female; 40% were born before 1980, 40% were born between 1980 and 1990, and 20% were born between 1990-1999 (Difallah, Catasta, Demartini, Ipeirotis, & Cudré-Mauroux, 2015; Ipeirotis, 2010).

Procedure

Participants learned by trial and error the average value of images from two categories (indoor or outdoor scenes) in two rooms defined by different background colors (see Figure 1.1). In each room, one type of scene was worth 40¢ on average (low-value category) and the other worth 60¢ (high-value category). The average values of the categories were matched across rooms, but the reward variance of the high-risk room was more than double that of the low-risk room (high-risk $\sigma = 34.25$, low-risk $\sigma = 15.49$). The order of the rooms (high-risk and low-risk) was randomized across participants. In an instruction phase, participants were explicitly told (through written instructions; see supplementary material) that in each room one scene category is worth more than the other (a 'winning' category) and were asked to indicate the winner after viewing all images in a room. They were not told the reward distributions of the rooms, nor that the rooms would have different levels of variance. In addition, to motivate participants to pay attention to individual scenes and their outcomes, participants were told that later in the experiment they would have the opportunity to choose between these same scenes and receive the rewards associated with them as per their choices.

After the two learning blocks (one high-risk and one low-risk), participants completed a risk attitude questionnaire (DOSPERT; Weber, Blais, & Betz, 2002) that served to create a 5-10 minute delay between learning and memory tests. Participants then completed a surprise item-recognition task (i.e., participants were never told that their memory for scenes would be tested, apart from instructions about the choice task as detailed above), as well as a sequence memory task. After the memory tests, participants made choices between previously seen images.

Learning: On each trial, participants were shown a trial-unique image (either an indoor or outdoor scene) for 2 seconds. Participants then had up to 5 seconds to estimate how much that type of scene is worth on average in that room (from 1 to 100 cents). In other words, participants were asked to provide their estimate of the average, or expected value, of the scene category based on the previous (variable) outcomes they had experienced from that scene category within the room. The scene was then presented again for 3 seconds along with its associated reward (see Figure 1.1A). In the instructions, participants had been told that although trial-unique images can take on different rewards, each scene category had a stable mean reward, and on average one scene category was worth more than the other. Note that participants were not asked to estimate the exact outcome they would receive on that trial, but instead were estimating the average expected reward from that scene category. Accordingly, participants had also been told that their payment was not contingent on how accurate their guesses were relative to the reward on that trial. Instead, their payment was solely determined by the rewards they received, to ensure that rewards were meaningful for the participant. This task structure was chosen to ensure that participants would continue to experience prediction errors on each trial (i.e., for individual scenes) even after correctly estimating the expected values of the categories, as is commonly done in reinforcement learning tasks (e.g., Niv, Edlund, Dayan, & O'Doherty, 2012).

There were 16 trials in each room (8 outdoor and 8 indoor). Rewards were 20¢, 40¢, 80¢, 100¢ (twice each) for the high-risk–high-value category, 0¢, 20¢, 60¢, 80¢ for the high-risk–low-value category, 45¢, 55¢, 65¢, 75¢ for the low-risk–high-value category and 25¢, 35¢, 45¢, 55¢ for the low-risk–low-value category. All participants experienced the same sequence of rewards within each room, with the order of the rooms randomized.

Memory: After completing the risk questionnaire, participants were presented with a surprise recognition memory test in which they were asked whether different scenes were old or new (see Figure 1.1B) as well as their confidence for that judgment (from 1 'guessing' to 4 'completely certain'). There were 32 test trials, including 16 old images (8 from each room) and 16 foils. Participants were then asked to sequence 8 pairs of previously seen scenes (which were not included in the recognition memory test) by answering 'which did you see first?' and by estimating how many trials apart the images had been from each other. Each pair belonged to either the low (4 pairs) or high-risk room (4 pairs).

Choice: In the last phase of the experiment, to verify that participants had encoded and remembered the individual outcomes associated with different scenes, participants were asked to choose between pairs of previously seen scenes for a chance to receive their associated reward again (see Figure 1.1C). The pairs varied in either belonging to the same room or different rooms and some were matched for reward and/or average scene value in order to test for the effects of factors such as risk context on choice preference. The choices were presented without feedback.

Statistical analysis: Analyses were conducted using paired t-tests, repeated measures ANOVAs, and generalized linear mixed-effects models (lme4 package in R; Bates et al., 2015). All results reported below (t-tests and ANOVAs) were confirmed using linear or generalized mixed-effects models treating participant as a random effect (for both the intercept and slope of the fixed effect in question). We note that in all experiments, our results held when controlling for the between-subjects variable of room order (for brevity, we only explicitly report these results in Experiment 1, see below).

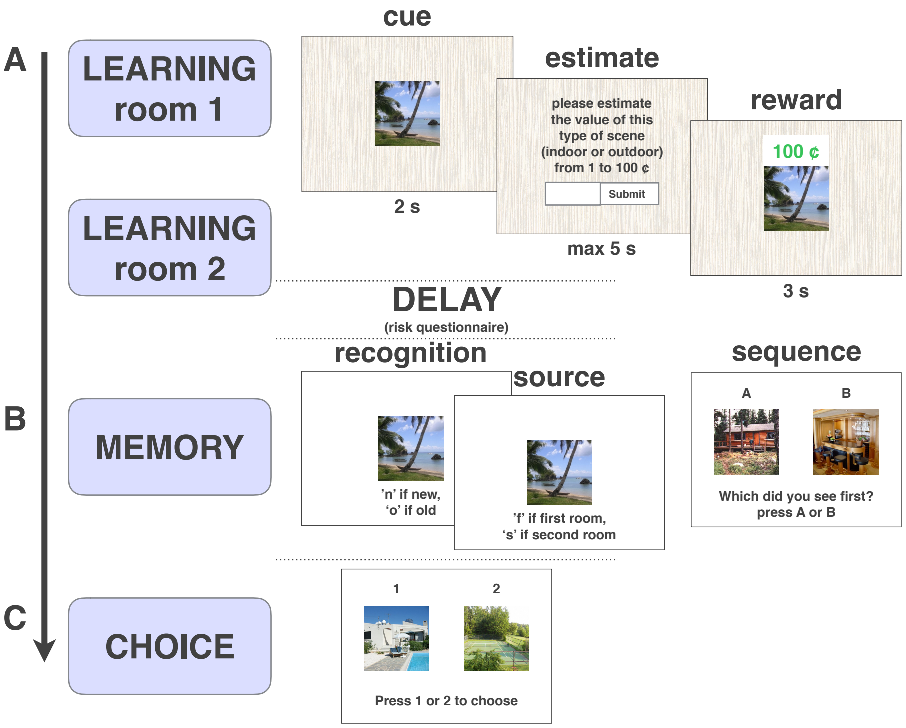


Figure 1.1: Task Design. A: Example learning trial. On each trial, participants were shown an image ('cue'), and were asked to estimate how much on average that type of scene (indoor or outdoor) was worth ('estimate'). They then saw the image again with a monetary outcome ('reward'). Each image appeared on one trial only. B: Memory tests. Participants completed item recognition, source (Exp. 2,3) and sequence memory tasks. C: Choice task. Participants chose between previously seen images that were matched for reward outcome, risk context, and/or scene category value.

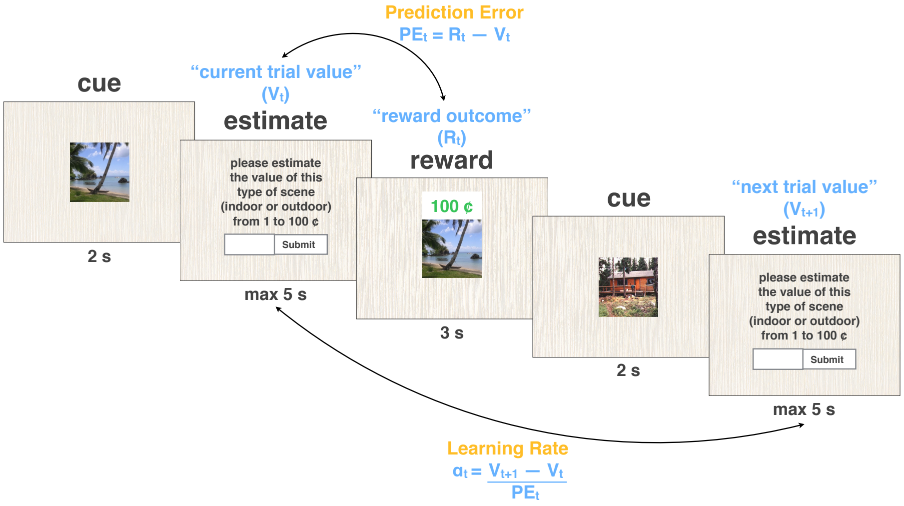


Figure 1.2: Schematic of prediction error (PE) and learning rate (α) calculation for two consecutive trials that involve the same scene category, in the learning phase of the experiment. Based on the learning equation $V_{t+1} = V_t + \alpha_t * PE_t$, we calculated the trial-by-trial learning rate as $(V_{t+1} - V_t)/PE_t$. Note that all components of this equation are measured explicitly: V_t and V_{t+1} are two consecutive estimates of the value of a scene from a single category (e.g., outdoor scenes), and the prediction error on trial t is the difference between the reward given on that trial, and the participants' estimate of the value of scene on the same trial. We assume here that separate values are learned and updated for each of the scene categories.

Results

Learning

Participants learned the average values of the high- and low-value categories better in the low-risk than in the high-risk room, as assessed by the deviation of their value estimates from the true averages of the scene categories ($t(163) = 14.52, p < 0.001$; Figure 1.3A). We then calculated, for every scene, the prediction error (PE_t) associated with that scene by subtracting participants' value estimates (V_t) from the reward outcome they observed (R_t; see Figure 1.2). This showed that, as we had planned, there were more high-magnitude prediction errors in the high-risk room as compared to the low-risk room ($t(163) = 36.77, p < 0.001$, within-subject comparison of average absolute prediction errors between the two rooms; Figure 1.3B).

Moreover, there was an interaction between risk and scene category such that participants overestimated the value of low-value scene category (resulting in negative prediction errors, on average) and underestimated the value of high-value scene category (resulting in positive prediction errors, on average) to a greater extent in the high-risk room than in the low-risk room ($F(1,163) = 141.2, p < 0.001$ for a within-subject interaction of the effects of room and scene category on the average signed prediction error; Figure 1.3C). This demonstrates more difficulty in separating the values of the categories in the high-risk room, consistent with previous findings showing that when people estimate the means of two largely overlapping distributions, they tend to average across the two distributions, thereby grouping them into one category instead of separating them into two (Gershman & Niv, 2013). Despite greater difficulty in separating the values of the high and low value categories within the high-risk room, most participants correctly guessed the 'winner', or the high-value scene category, within both the high-risk (88%) and the low-risk (89%) rooms.

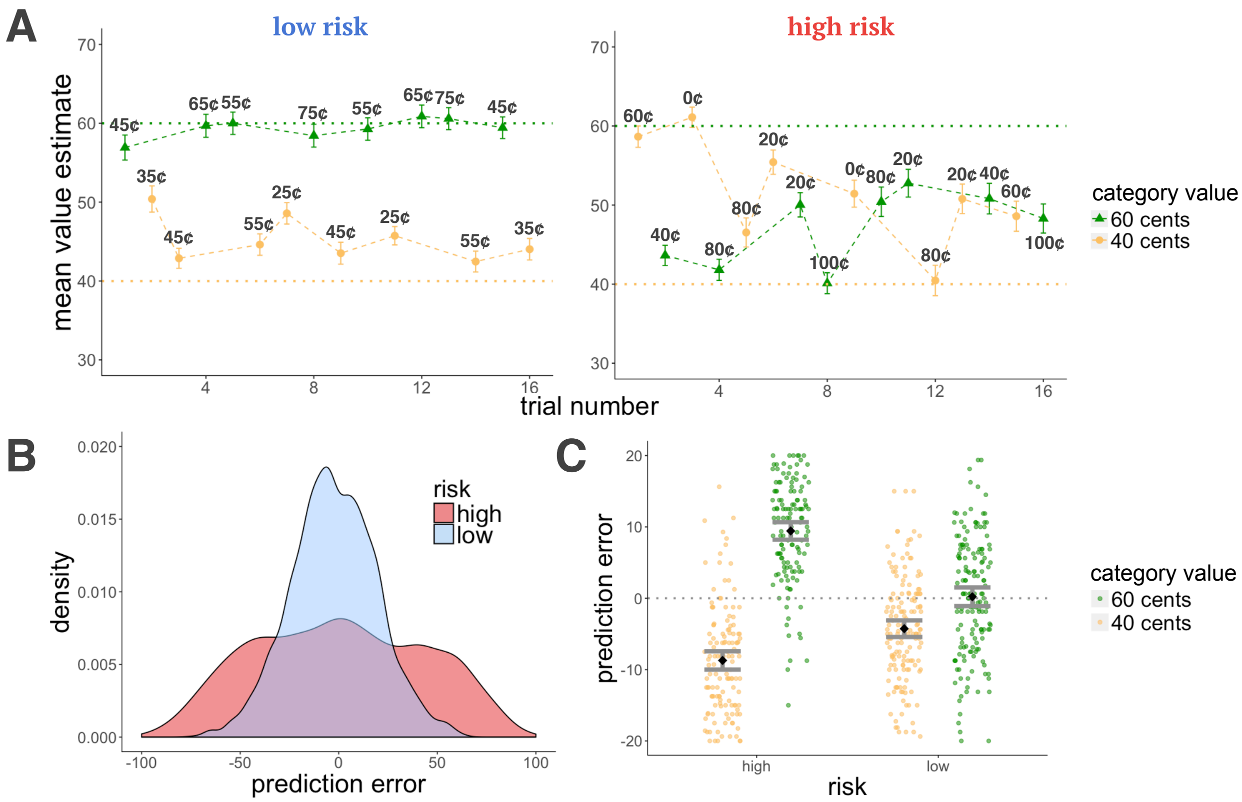


Figure 1.3: Experiment 1, learning results. A: Average estimates for the high and low-value categories as a function of trial number for the high and low-risk rooms. Participants learned better in the low-risk room, indicated by the proximity of their guesses to the true values of the scenes (dashed horizontal lines). Cent values represent the outcome participants received on that trial (after entering their value estimate). B: Density plot of prediction errors (PE_t) in each room. There were more high-magnitude prediction errors in the high-risk in comparison to the low-risk room. C: There was an interaction for positive and negative prediction errors between risk context and category value, such that participants overestimated the value of the low-value category and underestimated the value of the high-value category to a greater extent in the high-risk room. Error bars represent standard error of the mean.

Memory by risk and prediction error

We found that items within the high-risk room were recognized better than items within the low-risk room ($z = 2.37, p = 0.02, \beta = 0.31$; Figure 1.4A). To test the effect of reward prediction errors on item-recognition memory, we ran two separate mixed-effects logistic regression models of memory accuracy, one testing for the effect of signed and the other the effect of unsigned (absolute) prediction errors on recognition memory. Both models also included a risk-level regressor to test for the effects of risk and prediction error separately, and treated participants as a random effect. We did not find signed prediction errors to influence recognition memory beyond the effect of risk (signed prediction error (PE_t): $z = 0.71, p =$ n.s., $\beta = 0.04$; risk: $z = 2.29, p = 0.02, \beta = 0.30$). Instead, we found that larger prediction errors enhanced memory regardless of the sign of the prediction error, which also explained the modulation of memory by risk (absolute prediction error ($|PE_t|$): $z = 3.36, p < 0.001, \beta = 0.23$; risk: $z = 0.9, p =$ n.s., $\beta = 0.10$; Figure 1.4B).

We ran two subsequent models testing for confounds, one including the effect of value estimates and the other the actual reward outcomes associated with the items, along with the effect of absolute prediction errors. Absolute prediction error had a significant effect on recognition memory when controlling for reward outcome ($|PE_t|$: $z = 3.94, p < 0.001, \beta = 0.26$; R_t: $z = 0.45, p =$ n.s., $\beta = 0.02$) and value estimates ($|PE_t|$: $z = 3.93, p < 0.001, \beta = 0.26$; V_t: $z = -0.09, p =$ n.s., $\beta = -0.005$). This effect also held when modeling recognition memory for items in the high and low-risk rooms separately (high-risk: $z = 1.90, p = 0.05, \beta = 0.18$; low-risk: $z = 2.17, p = 0.03, \beta = 0.24$), and in a model of the effects of absolute prediction errors on recognition memory that controlled for room order ($|PE_t|$: $z = 3.90, p < 0.001, \beta = 0.25$; room order: $z = 1.95, p = 0.05, \beta = 0.33$). Although room order itself did affect recognition memory (participants who experienced the low-risk room first showed better memory accuracy overall), all of our main findings (including learning rate) held when controlling for this effect.

Reward prediction errors therefore affected recognition memory, such that larger deviations from one's predictions, in any direction, enhanced memory for items. Finally, we tested for the effect of risk on sequence memory (the correct ordering of two images seen during learning) and found no difference in sequence memory between pairs of images seen in the high and low-risk rooms ($z = 0.11, p =$ n.s., $\beta = 0.02$).

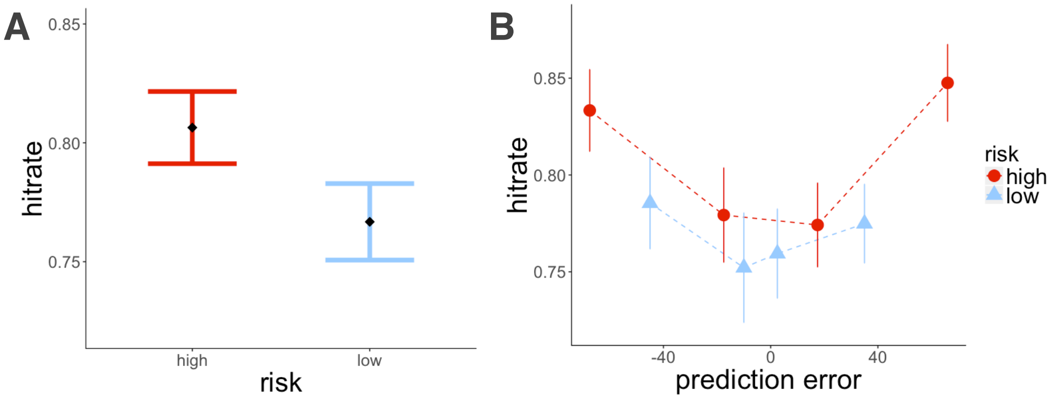


Figure 1.4: Experiment 1, recognition memory results. A: Recognition memory was better for items within the high-risk room. B: There was better recognition memory for items associated with a higher absolute prediction error. Item memory was binned by the quartile values of prediction errors within each risk room. Each dot represents the average value within that quartile. Error bars represent standard error of the mean.

Learning rate by risk and prediction error

We also examined the effects of risk and prediction errors on the reward learning process itself. For this we calculated a trial-by-trial learning rate α_t as the proportion of the current prediction error, $PE_t = R_t - V_t$, that was applied to update the value for the next encounter of the same type of scene, V_{t+1} (see Figure 1.2 for schematic representing learning rate calculation). That is, we derived the trial-specific learning rate directly from the standard reinforcement-learning update equation $V_{t+1} = V_t + \alpha_t(R_t - V_t)$, as:

$$\alpha_t = \frac{V_{t+1} - V_t}{R_t - V_t}.$$

In agreement with recent findings (Diederen et al., 2016), we found that average learning rate was higher in the low-risk room than in the high-risk room ($t(163) = 3.37, p < 0.001$ within-subjects test; Figure 1.5A). Moreover, higher absolute prediction errors increased trial-by-trial learning rates (α_t) above and beyond the effect of risk (mixed-effects linear model, effect of absolute prediction error: $t = 3.30, p = 0.001, \beta = 0.07$; risk: $t = 4.67, p < 0.001, \beta = 0.16$; Figure 1.5B). We did not find participant room order to influence learning rate ($t = 0.31, p =$ n.s., $\beta = -0.03$). These results show that larger absolute prediction errors enhance value updating, and further, that learning rates adapt to the reward variance such that there is greater sensitivity to prediction errors in a lower-risk environment.

We next ran a mixed-effects regression model to test whether trial-by-trial learning rates predicted recognition memory for scenes at test. Controlling for absolute prediction error, we did not find that learning rate on trial t predicted memory on that same trial (α_t: $z = 0.85, p =$ n.s., $\beta = 0.08$; $|PE_t|$: $z = 3.42, p < 0.001, \beta = 0.20$), nor on the subsequent trial, (effect of α_{t-1}) on recognition memory for the scene on trial t: $z = 0.56, p =$ n.s., $\beta = 0.05$; $|PE_t|$: $z = 3.06, p = 0.002, \beta = 0.19$, where t enumerates over trials within a room). This demonstrates that increases in learning rate were not correlated with better (or worse) memory, even though both learning rate and recognition memory were enhanced by larger prediction errors.

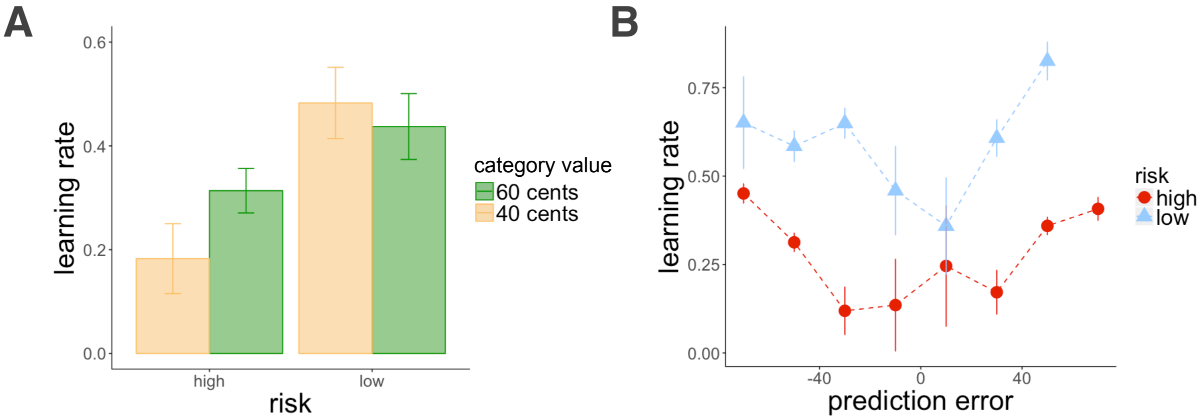


Figure 1.5: Experiment 1, learning rate results. A: Learning rate was higher in the low-risk context. Average learning rate plotted by risk context and category value. B: Both absolute prediction errors and a low-risk context increased learning rate. Learning rates were binned by prediction errors that occurred on the same trial (each dot represents the average prediction error within the binned range). Error bars represent standard error of the mean.

Choice by reward and value difference

Finally, in a manipulation test, participants were asked to make choices between pairs of previously-seen scenes. Choices between scenes with different reward outcomes served to test whether participants encoded the rewards associated with the images. Participants chose the image associated with the larger outcome more often (mixed-effects logistic regression model predicting choice based on outcome: $z = 6.40, p < 0.001, \beta = 0.54$), suggesting that they did indeed encode and remember the rewards associated with the scenes.

Some choices were between items that were associated with the same outcome feedback. Here we sought to test whether features of the environment such as the risk context biased participants away from indifference. We did not find risk level, whether the scene was from the low rewarding or high rewarding category, or the difference in absolute prediction error between the images, to additionally influence choice preference. We instead found that participants were more likely to choose the scene that they had initially guessed a higher value for ($z = 3.74, p < 0.001, \beta = 0.01$). We additionally found that even when the two options had led to different reward outcomes, the difference in initial value estimates for the scene was a significant predictor of choice, above and beyond the difference in actual reward outcome (value estimate difference: $z = 2.27, p = 0.02, \beta = 0.16$; reward difference: $z = 7.25$, $p < 0.001, \beta = 0.52$). This suggests that participants remembered not only the outcomes for different scenes, but also their initial estimates.

Discussion

In Experiment 1, we showed that the greater the magnitude of the prediction error experienced during value learning, the more likely participants were to recognize items associated with those prediction errors. We also demonstrated that both risk context and absolute prediction errors influenced the extent to which people updated values for the scene categories, i.e. their item-by-item learning rate fluctuated according to prediction errors and was influenced by context. In particular, learning rate was higher in the low-risk environment, suggesting greater sensitivity to prediction errors when the variance of the environment was lower. Further, in both contexts, higher absolute prediction errors increased learning rate. Although absolute prediction errors enhanced both recognition memory and learning rate, we did not find learning rate to predict recognition memory, suggesting that absolute prediction errors affect learning and memory through distinct mechanisms.

1.2 EXPERIMENT 2

In Experiment 2, we allowed for more learning in both rooms, which posed stronger memory demands. We also tested for other types of episodic memory. Notably, different from standard reinforcement-learning paradigms, Experiment 1 involved only 16 trials of learning in each context, 8 for each category. The initial phase of learning, which we were effectively testing, is characterized by increased prediction errors and uncertainty relative to later learning, which might affect the relationship between prediction errors and episodic memory. Additionally, participants in Experiment 1 all experienced the same reward sequence, which inadvertently introduced regularities in the learning curves that could have also influenced initial learning and memory results. Finally, in this relatively short experiment, average recognition memory performance was near ceiling (A' = 0.90). In Experiment 2, we therefore sought to replicate the results of Experiment 1 while increasing the number of learning and memory trials and randomizing reward sequence. With more trials, we were also able to test for sequence memory for items that were presented further apart in time, and we included a measure of source memory (i.e., which room the item belonged to)—a marker of episodic memory—for the context of the probed item.

Method

Participants

Two hundred participants initiated an online task run on Amazon Mechanical Turk, and 148 completed the task. Following the same protocol as in Experiment 1, twelve participants were excluded from the analysis leading to a final sample of 136 participants.

Procedure

The procedure was the same as in Experiment 1 but with some changes to learning, memory and choice. As in Experiment 1, rewards had a mean of 60¢ for the high-value category and 40¢ for the low-value category (high-risk–high-value scenes: 20¢, 40¢, 60¢, 80¢, 100¢; high-risk–low-value scenes: 0¢, 20¢, 40¢, 60¢, 80¢; low-risk–high-value scenes: 40¢, 50¢, 60¢, 70¢, 80¢; low-risk–low-value scenes: 20¢, 30¢, 40¢, 50¢, 60¢). However, we increased the number of learning trials from 16 to 30 trials per room, and we pseudo-randomized the reward sequence such that the rewards were drawn at random and were sampled three times without replacement.

During the item memory test, we also asked participants to indicate whether items identified as 'old' belonged to the first or second room (see Figure 1.1B), to measure source memory. Additionally, given that sequence memory improves with greater distance between events (DuBrow & Davachi, 2013), here we asked participants to order items that were as far as 13-14 trials apart, in contrast to the maximum of 8 trials apart in Experiment 1. Finally, satisfied by the manipulation check in the choice tasks in Experiment 1, we asked participants to choose only between pairs of scenes matched for reward outcome.

Results

Learning

As in Experiment 1, participants learned better in the low-risk than in the high-risk room (assessed by the average deviation of participants' value estimates from the true means of the category values; $t(135) = 13.11, p < 0.001$; Figure 1.6A). They experienced larger absolute prediction errors in the high-risk room ($t(135) = 39.65, p < 0.001$; Figure 1.6B), and there was again an interaction between risk and scene category value such that in the high-risk room, participants overestimated the value of the low-value scene category and underestimated the value of the high-value scene category to a greater extent than in the low-risk room ($F(1,135) = 77.5, p < 0.001$; interaction of the effects of room and category on average prediction error experienced; Figure 1.6C). Again, participants guessed the high-value scene category at the end of each room equally well in the high-risk (90%) and low-risk (89%) rooms.

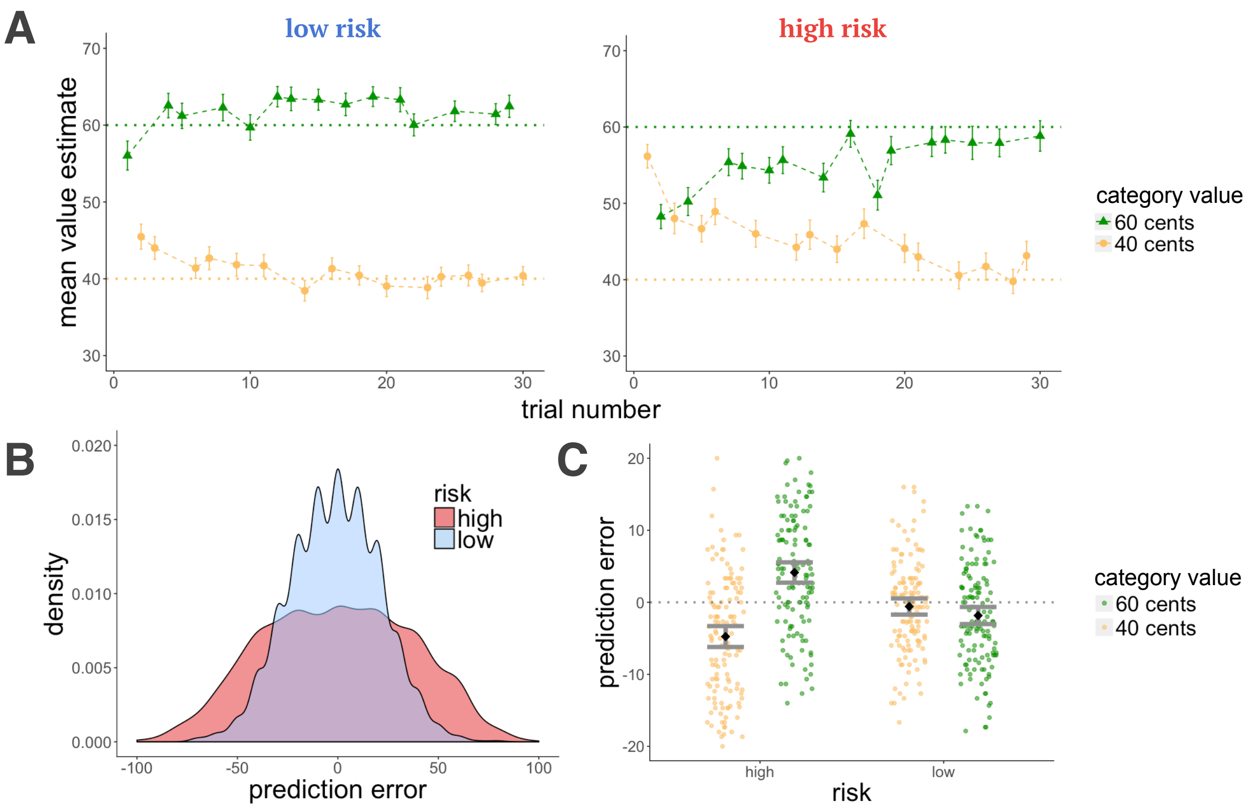


Figure 1.6: Experiment 2, learning results. A: Average estimates for the high and low-value categories as a function of trial number for the high and low-risk rooms. Participants learned better in the low-risk room, indicated by the proximity of their guesses to the true values of the scenes (dashed horizontal lines). B: Density plot of prediction errors (PE_t) in each room. There were more high-magnitude prediction errors in the high-risk in comparison to the low-risk room. C: There was an interaction for positive and negative prediction errors between risk context and category value, such that participants overestimated the value of the low-value category and underestimated the value of the high-value category to a greater extent in the high-risk room. Error bars represent standard error of the mean.

Memory by risk and prediction error

By increasing the number of learning and memory trials, we significantly reduced average recognition memory performance from Experiment 1 (A' = 0.86, $t(275.23) = 3.04$, $p = 0.003$ when comparing overall memory performance between Experiment 1 and 2). We nevertheless replicated the main results of Experiment 1: items from the high-risk room were better recognized than items from the low-risk room ($z = 2.51$, $p = 0.01$, $\beta = 0.19$ when testing for the effect of risk on item-recognition memory; Figure 1.7A). In a separate model, higher absolute prediction errors enhanced recognition memory for scenes, while again explaining the effect of risk ($|PE_t|$: $z = 3.44$, $p < 0.001$, $\beta = 0.16$; risk: $z = 1.76$, $p = 0.08$, $\beta = 0.14$, Figure 1.7B). Like in Experiment 1, in subsequent models testing for potential confounds, this effect was significant when controlling for the outcomes associated with the items ($|PE_t|$: z $= 4.14$, $p < 0.001$, $\beta = 0.18$; outcome R_t: $z = -1.71$, $p =$ n.s., $\beta = -0.06$) as well as for the value estimate for the scene category ($|PE_t|$: $z = 4.15$, $p < 0.001$, $\beta = 0.19$; estimate V_t: $z = -1.16$, $p =$ n.s., $\beta = -0.04$).

In addition, for the scenes correctly identified as old, we found better source memory for scenes from the high-risk room ($z = 2.05$, $p = 0.04$, $\beta = 0.25$ in a mixed-effects logistic regression model testing for the effect of risk on source memory; Figure 1.7C). This effect was not modulated by absolute prediction error. Rather, it was a context effect: the source of a recognized image was better remembered if that item was seen in the high-risk room (absolute prediction errors: $z = -0.60$, $p =$ n.s., $\beta = -0.03$; risk: $z = 2.17$, $p = 0.03$, $\beta = 0.27$). To verify that participants were not simply attributing remembered items to the high-risk context, we looked at the proportion of high-risk source judgments for recognition hits and false alarms separately. We did not find a greater proportion of high-risk source judgments for false alarms, indicating that participants were not biased to report that remembered items belonged to a high-risk context (for high-risk hits: mean = 0.57, standard error = 0.02; for false alarms: mean = 0.49, standard error = 0.04; chance response is 0.50).

Participants also exhibited better sequence memory for pairs from the high-risk room ($z = 2.70$, $p = 0.007$, $\beta = 0.56$ in a mixed-effects logistic regression model testing for the effect of risk on sequence memory; Figure 1.7D). Although we did not see this effect in Experiment 1, the longer training in Experiment 2 allowed us to test pairs that were more distant from each other (the most distant items were 13 and 14 trials apart). Indeed, in a model additionally testing for the effect of distance between tested pairs, greater distance predicted better sequence memory, controlling for risk (distance: $z = 1.92$, $p =$

0.05, $\beta = 0.39$; risk: $z = 2.70$, $p = 0.006$, $\beta = 0.56$). We therefore replicated our original results and further showed that other forms of episodic memory—source and sequence memory—were also enhanced in a high-risk context.

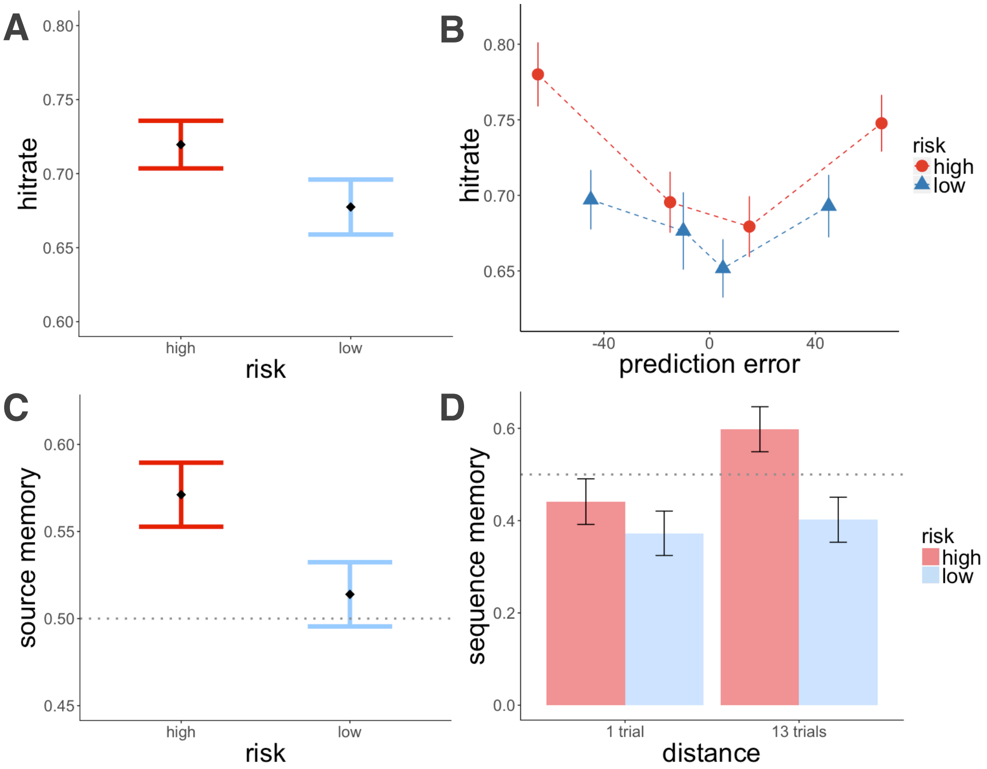


Figure 1.7: Experiment 2, memory results. A: Recognition memory was better for items within the high-risk context. B: Absolute prediction errors enhanced recognition memory for the scenes. Item memory was binned by the quartile values of prediction errors within each risk room, each dot represents the average value within that quartile. C: For correctly remembered items, source memory was better for items within the high-risk context. D: A high-risk context and distance between items (number of trials between pairs) increased sequence memory. Error bars represent the standard error of the mean.

Learning rate by risk and prediction error

We replicated the results of Experiment 1 with respect to learning rates as well: participants had higher learning rates for the low-risk relative to the high-risk room, and higher absolute prediction errors additionally increased learning rates in a mixed-effects regression model testing for the effect of risk and absolute prediction error on learning rate (absolute prediction error: $t = 5.12$, $p < 0.001$, $\beta = 0.09$; risk: $t = 7.01$, $p < 0.001$, $\beta = 0.18$; Figure 1.8A-B). Controlling for absolute prediction error, we again did not find learning rate to predict recognition memory on the current trial (α_t: $z = -0.29$, $p =$ n.s., $\beta = -0.01$; $|PE_t|$: $z = 4.44$, $p < 0.001$, $\beta = 0.20$), nor the subsequent trial (α_{t-1}: $z = 0.68$, $p =$ n.s., $\beta = 0.03$;

$|PE_t|$: $z = 3.53, p < 0.001, \beta = 0.17$).

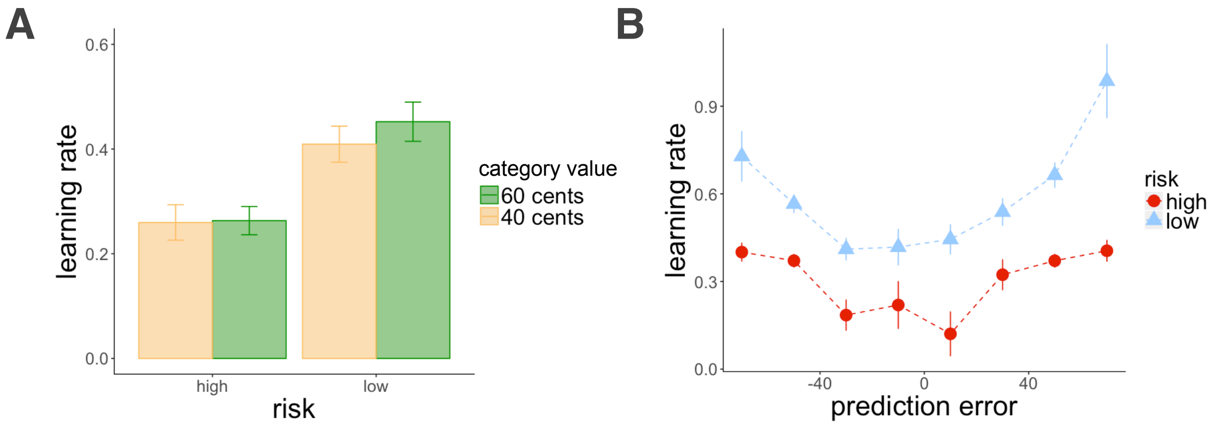


Figure 1.8: Experiment 2, learning rate results. A: Learning rate was higher in the low-risk context. Average learning rate plotted by risk context and category value. B: Both absolute prediction errors and a low-risk context increased learning rate. Learning rates were binned by prediction errors that occurred on the same trial (each dot represents the average prediction error within the binned range). Error bars represent standard error of the mean.

Choice by reward and value difference

In this experiment, all choices were between images with matched reward outcomes. We replicated the results of Experiment 1 such that choice was predicted by the difference in participants' initial value estimates for the scenes ($z = 2.78, p = 0.005, \beta = 0.18$). In particular, even in this better-powered test (12 choice trials as compared to 4 with matched outcomes in Experiment 1), there was no evidence for preference for images from one risk context versus the other ($z = -1.56$, $p =$ n.s., $\beta = -0.08$).

Discussion

In Experiment 2, we doubled the number of training trials and replicated the results of Experiment 1, showing that large prediction errors increase learning rate and improve recognition memory, but that higher learning rates do not predict better item recognition. In fact, like in Experiment 1, learning rates were higher in the low-risk room, but item recognition was better in the high-risk room. Moreover, in this experiment, we demonstrated additional risk-context effects on episodic memory by showing better sequence and source memory for items that were encountered in the high-risk learning context. These results were separate from the effect of absolute prediction errors, but perhaps point to general memory enhancement for events occurring in a putatively more arousing environment.

1.3 EXPERIMENT 3

A possible confound of the effects of risk on memory and learning in Experiments 1 and 2 is that there was higher overlap between the outcomes for the two categories in the high-risk context as compared to the low-risk context. The distributions of outcomes for the indoor and outdoor scenes shared values from 20¢ to 80¢ (Exp. 1 & 2) in the high-risk room, but only 45¢ to 55¢ (Exp. 1) and 40¢ to 60¢ (Exp. 2) in the low-risk room. This greater overlap in the high-risk context could have made learning more difficult in comparison to the low-risk room, and therefore influenced the effects of absolute prediction error on subsequent memory. To test for this possibility, in Experiment 3 we made the learning conditions in the two rooms more similar by eliminating any overlap between the outcomes of the two scene categories.

Method

Participants

We conducted a simulation-based power analysis of the effect of absolute prediction errors on item-recognition memory. This revealed that we would have sufficient power (80% probability) to replicate the results of Experiments 1 and 2 with as few as 55 participants. As a result, we had 100 participants initiate the study, of which 86 completed the task. Three participants were excluded based on our exclusion criteria (see Experiment 1) leaving a final sample of 83 participants.

Procedure

We followed the same procedure as in Experiment 2 but changed the rewards such that they had a mean of 80¢ for the high-value category and 20¢ for the low-value category, and there was no overlap between the outcomes for scenes from the two categories (high-risk–high-value scenes: 60¢, 70¢, 80¢, 90¢, 100¢; high-risk–low-value scenes: 0¢, 10¢, 20¢, 30¢, 40¢; low-risk–high-value scenes: 70¢, 75¢, 80¢, 85¢, 90¢; low-risk–low-value scenes: 10¢, 15¢, 20¢, 25¢, 30¢).

Results

Learning

As in Experiment 1 and 2, participants learned better in the low-risk than in the high-risk room ($t(82) = 6.28, p < 0.001$ in a paired t test comparing the average deviation of estimates from the true means of the categories across rooms; Figure 1.9A). However, learning in the two rooms was more similar here than in Experiment 2, as assessed by first computing the difference in learning (average deviation of estimates from the true means of the scene categories) between the high and low-risk rooms for each participant, and then comparing this value between participants in Experiments 2 and 3 ($t(148.98) = 1.84, p = 0.03$). The range of prediction errors in the two rooms was also more similar in comparison to Experiment 1 and 2 (Figure 1.9B), allowing us to better assess the effects of risk context on learning and memory, when controlling for prediction errors (see below). As in previous experiments, there was an interaction between risk and scene category such that participants overestimated the low-value category and underestimated the high-value category more in the high-risk than in the low-risk room, ($F(1,82) = 23.02, p < 0.001$; Figure 1.9C). Nonetheless, participants correctly guessed the high-value category equally well (and at a higher proportion than in Experiment 1 and 2) in the high-risk (95%) and low-risk (96%) rooms.

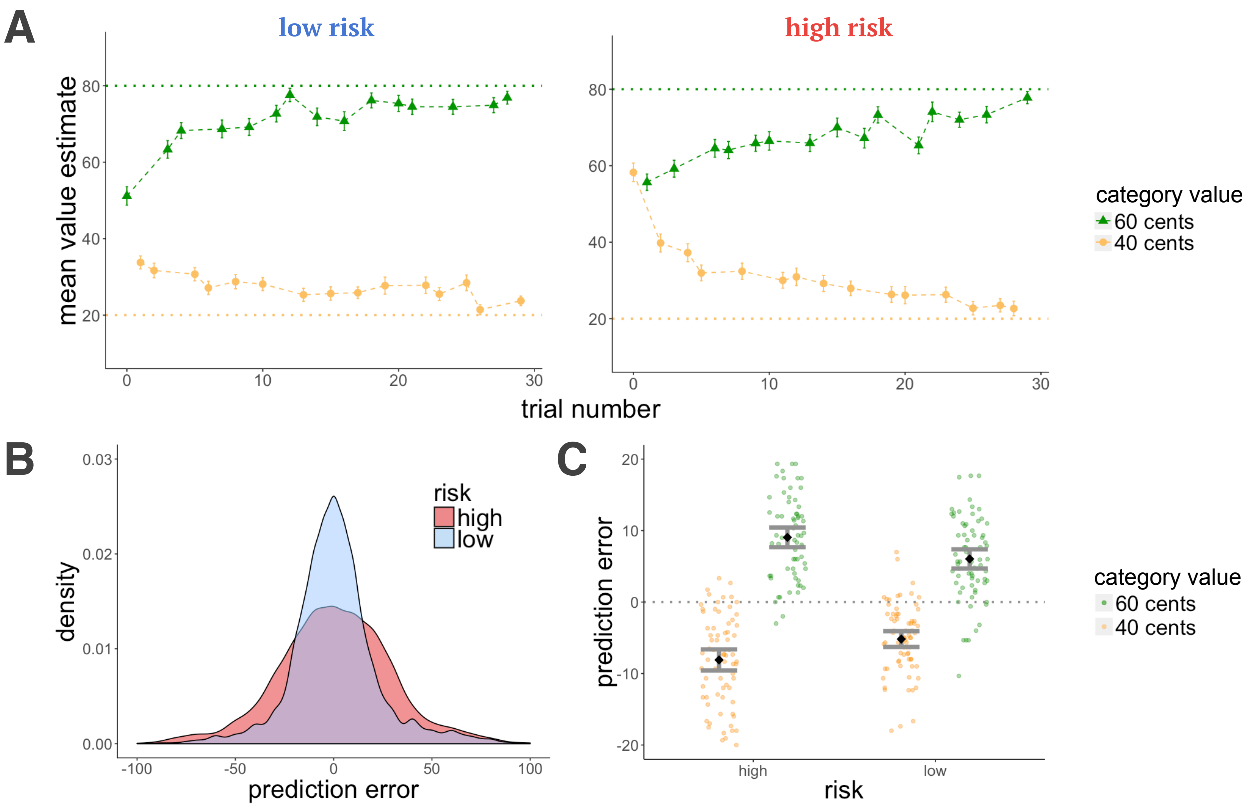


Figure 1.9: Experiment 3, learning results. A: Average estimates for the high and low-value categories as a function of trial number, separately for the high-risk and low-risk rooms. Participants learned better in the low-risk room (although the difference in learning between risk rooms was smaller than in Exp. 1 & 2). B: Density plot of experienced prediction errors (PE_t) in each room. Compared to Exp 1 & 2, there were higher-magnitude prediction errors in the low-risk room, making the range of prediction errors more similar between rooms. C: Prediction errors show an interaction between risk context and category value, such that participants overestimated the value of the low-value category and underestimated the value of the high-value category to a greater extent in the high-risk room. Error bars represent the standard error of the mean.

Memory by risk and prediction error

We replicated the results of Experiments 1 and 2, and further found separate effects of context and unsigned prediction error on recognition memory. A high-risk context and larger absolute prediction errors enhanced recognition memory for scenes, even with both predictors in the same model, indicating independent effects ($|PE_t|$: $z = 2.24$, $p = 0.02$, $\beta = 0.12$; risk: $z = 2.58$, $p = 0.009$, $\beta = 0.24$, Figure 1.10A-B). This effect was again significant when controlling for reward outcome ($|PE_t|$: $z = 2.72$, $p = 0.007$, $\beta = 0.15$; R_t: $z = -0.38$, $p =$ n.s., $\beta = -0.02$) and value estimates ($|PE_t|$: $z = 2.70$, $p = 0.007$, $\beta = 0.15$; V_t: $z = -0.74$, $p =$ n.s., $\beta = -0.03$). Similar to Experiment 2, we again found better sequence memory for items within the high-risk context, while controlling for the effect of distance (risk: $z = 2.47$, $p = 0.01$, $\beta = 0.57$; distance: $z = 2.36$, $p = 0.02$, $\beta = 0.55$). For source memory, we did not have the power to detect the effect in Experiment 2, and this difference was not statistically significant although it was in the same direction.

It is worth noting here that there was a stronger effect of context in modulating recognition memory than in Experiments 1 and 2 (the context effect remained when controlling for absolute prediction errors, unlike in Experiments 1 and 2). That is, when learning was more similar in the two rooms, an independent effect of risk in increasing recognition memory became apparent. One possible explanation for this finding is that memory-boosting effects of reward prediction errors might "spill over" to adjacent trials, enhancing memory for those items as well. To test for these "spill over" effects in the high-risk context, we measured whether immediately previous and subsequent absolute prediction errors proactively or retroactively strengthened recognition memory for a scene, while controlling for the absolute prediction error experienced for that particular scene. We ran two mixed-effects logistic-regression models testing for the effect of adjacent absolute prediction errors (one for previous and one for subsequent prediction error) on recognition memory. We did not find any effect of adjacent prediction errors ($|PE_{t-1}|$: $z = -1.71$, $p =$ n.s., $\beta = -0.13$; $|PE_{t+1}|$: $z = -0.93$, $p =$ n.s., $\beta = -0.08$), suggesting that the memory-enhancing effect of the high-risk context may be due to general enhanced memory for items experienced in a high-risk, and potentially more arousing, environment.

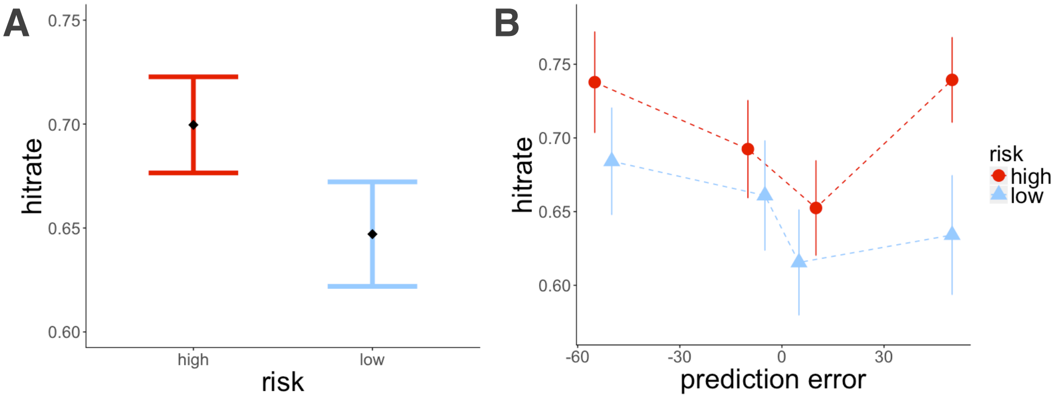


Figure 1.10: Experiment 3, recognition memory results. A: Recognition memory was better for scenes that were encountered in the high-risk context. B: Both absolute prediction errors and a high-risk context independently enhanced recognition memory for scenes. Item memory was binned by the quartile values of prediction errors within each risk room. Each dot represents the average value within that quartile. Error bars represent the standard error of the mean.

Learning rate by risk and prediction error

As in Experiments 1 and 2, absolute prediction errors increased learning rates in both rooms, and there was a trend for higher learning rates in the low-risk room ($|PE_t|$: $t = 3.33$, $p < 0.001$, $\beta = 0.06$; risk: $t = 1.84$, $p = 0.06$, $\beta = 0.06$; Figure 1.11A-B). We again did not find learning rate for values to predict recognition memory for the scene on the current trial ($z = -0.26$, $p =$ n.s., $\beta = -0.01$), nor the subsequent trial ($z = -1.22$, $p =$ n.s., $\beta = -0.08$), while controlling for the effect of absolute prediction error on the current trial.

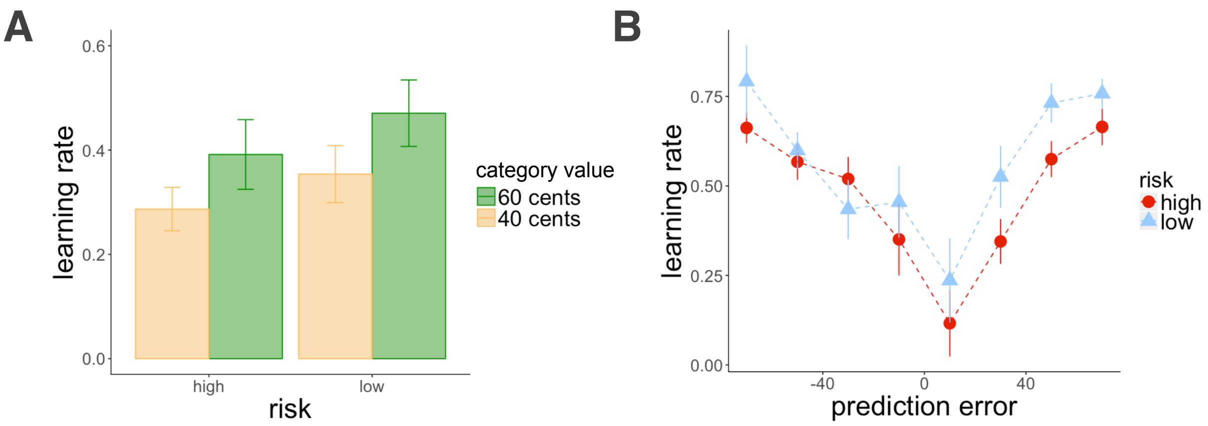


Figure 1.11: Experiment 3, learning rate results. A: There was a trend for higher average learning rates in the low-risk context. B: Absolute prediction errors increased learning rate. Learning rates were binned by prediction errors on the same trial (each dot represents the average prediction error within the binned range). Error bars represent standard error of the mean.

Choice by reward and value difference

As in Experiment 2, all choices (12 trials) were between scenes that had matched reward outcomes. Here too we replicated previous results, such that participants were more likely to choose the scene that they had initially guessed a higher value for ($z = 3.98$, $p < 0.001$, $\beta = 0.29$).

Discussion

In Experiment 3, we eliminated all overlap between the reward outcomes of the high and low-value categories in both rooms—a potential confound in Experiment 1 and 2—and replicated our previous results. Additionally, given the more similar range of prediction errors in the high and low-risk contexts, we were able to detect an independent effect of risk context on recognition memory. Improved recognition memory in the high-risk room, like the better source and sequence memory observed for high-risk events in Experiment 2, points to general memory enhancement for events experienced in an environment with greater reward variance.

1.4 General Discussion

Our aim was to determine how reward prediction errors influence episodic memory, above and beyond their known influence on learning. In Experiment 1, we demonstrated that unsigned, or absolute prediction errors enhanced recognition memory for a rewarding episode. That is, trial-unique scenes that were accompanied by a large reward prediction error, whether positive (receiving much more reward than expected) or negative (receiving much less reward than expected) were better recognized in a subsequent surprise recognition test. We additionally found that risk context and absolute prediction errors modulated the trial-by-trial learning rate by which participants used the rewards to update their estimate of the general worth of that category of scenes. In particular, learning rate was higher in a low-risk environment, and there was more learning from rewards that generated larger prediction errors. Notably, although large prediction errors increased learning from rewards on that specific trial, and enhanced memory for the scene in the trial, we did not find a trial-by-trial relationship between learning rate and memory accuracy. In fact, the high-risk context led to lower learning rates but better recognition memory on average, suggesting separate mechanisms underlying these two effects of prediction errors.

In Experiment 2, we increased the number of trials therefore allowing for more learning in each context, and placing more demands on memory. We replicated all the effects from Experiment 1, and further showed that source and sequence-memory were better for images encountered in the high-risk context. In Experiment 3, we eliminated a potential confound by equating learning difficulty in the high-risk and low-risk contexts, again reproducing the original results. This manipulation also resulted in a more similar range of prediction errors in both risk contexts, which uncovered a separate effect of risk on episodic memory, above and beyond that of absolute prediction errors.

Previous work has shown both a collaboration between learning and memory systems, such as boosting of memory for items experienced during reward anticipation (Adcock, Thangavel, Whitfield-

Gabrieli, Knutson, & Gabrieli, 2006) including oddball events (Murty & Adcock, 2014), as well as a competition between the systems, where the successful encoding of items experienced prior to reward outcome is thought to interfere with neural prediction errors (Wimmer, Braun, Daw, & Shohamy, 2014). Here, in all three experiments, we showed that incremental learning and episodic memory systems collaborate, as learning signals. Specifically, large reward prediction errors both increase learning rate for the value of the rewarding source and enhance memory for the scene that led to the prediction error. However, the fact that the effects of prediction errors on learning rate and episodic memory were uncorrelated suggests that these effects are mediated by somewhat separate neural mechanisms.

Although we only tested behavior, the impetus for our experiments were neurobiological accounts adjudicating between the effects of signed and unsigned reward prediction errors on memory. Neurally, reward prediction error modulation of dopamine signaling provides a strong putative link between trial-and-error learning and dopamine-induced plasticity in the hippocampus. Such an effect of (signed) dopaminergic prediction errors from the ventral tegmental area (VTA) to the hippocampus would have predicted an asymmetric effect on memory, such that memories benefit from a positive prediction error (signaled by an increase in dopaminergic firing from the VTA), but not a negative prediction error (signaled by decreased dopaminergic firing). Instead, we found that the absolute magnitude of reward prediction errors, regardless of the sign, enhanced memory. This mechanism perhaps explains the finding that extreme outcomes are recalled first, are perceived as having occurred more frequently, and increase preference for a risky option (Ludvig, Madan, & Spetch, 2014; Madan, Ludvig, & Spetch, 2014).

In our task, each outcome was sampled with equal probability (uniform distributions), meaning that extreme outcomes were not rare. However, the mnemonic effects that we identified could potentially also contribute to the well-demonstrated phenomenon of nonlinear responses to reward probability in choice and in the brain, characterized by the overweighting of low-probability events and the underweighting of high-probability ones (Hsu, Krajbich, Zhao, & Camerer, 2009; Kahneman & Tversky, 1979). In particular, large prediction errors due to the occurrence of rare events would mean that these events affect learning and memory disproportionately strongly. Similarly, the underweighting of very common events could arise from the rare cases in which the common event does not occur, giving rise to large and influential prediction errors. Our results suggest that these distortions of weighting would be especially prominent when episodic memory is used in performing the task.

The influence of unsigned reward prediction errors on recognition memory is also reminiscent of work demonstrating better memory for surprising feedback outside of reinforcement learning, such as a recent study showing improved encoding of unexpected paired associates (Greve, Cooper, Kaula, Anderson, & Henson, 2017). Another potentially related paradigm is the hypercorrection effect (Butterfield & Metcalfe, 2001), where high-confidence errors and low-confidence correct feedback (both potentially generating large prediction errors) lead to greater attentional capture and improved memory (Butterfield & Metcalfe, 2006).

Neuroscientific work has linked surprising feedback to increases in arousal and the noradrenergic locus coeruleus (LC; Clewett et al., 2014; Mather et al., 2015; Miendlarzewska, Bavelier, & Schwartz, 2016). Our finding that absolute prediction errors influenced subsequent memory is in line with a mechanism (also described in the Introduction) whereby the LC-norepinephrine system responds to salient (surprising) events, and dopamine co-released with norepinephrine from LC neurons strengthens hippocampal memories (Kempadoo et al., 2016; Takeuchi et al., 2016). This proposed mechanism would seem to imply that increases in learning rate (previously linked to norepinephrine release) and enhanced episodic memory (linked to dopamine release) should be correlated across trials, given the hypothesized common cause of LC activation. However, we found that increases in learning rate were uncorrelated with enhanced memory, suggesting that the actual mechanism may involve additional (or different) steps from the one described above.

In our task, learning rate not only increased with the magnitude of prediction error, but also changed with the riskiness of the environment. In line with our results, recent work shows that learning rate scales inversely with reward variance, with higher learning rates in lower variance contexts (Diederen & Schultz, 2015; Diederen et al., 2016). Greater sensitivity to the same magnitude prediction errors in a low versus a high-variance environment demonstrates adaptation to reward statistics, where in a low-risk context, even small prediction errors are more relevant to learning than they would be when there is greater reward variance. This heightened sensitivity to unexpected rewards in the low-risk environment, however, was not associated with improved episodic memory in any of our experiments. In fact, in Experiment 3, we found that memory was better for items experienced in the high-risk context, even when controlling for the magnitude of trial-by-trial reward prediction errors. The opposing effects of risk on learning rate and episodic memory again suggest distinct underlying mechanisms, in agreement with work characterizing learning and memory systems as separate and even antagonistic (Foerde, Braun, & Shohamy, 2012; Wimmer et al., 2014).

To explain the beneficial effect of high-risk environments on episodic memory, we hypothesized that better memory for large-prediction-error events could potentially "spill over" to surrounding items, in line with work showing that inducing an "encoding" state (such as through the presentation of novel items) introduces a lingering bias to encode subsequent items (Duncan & Shohamy, 2016; Duncan et al., 2012). These effects, however, did not explain how risk context modulated memory in our task, as we did not find prediction error events to additionally improve memory for adjacent items. Instead, we speculate that this context effect is due to improved encoding when in a putatively more aroused state, although future studies should more directly characterize the link between arousal and enhanced memory in risky environments.

Finally, we did not find effects of absolute prediction error or risk context on preferences in a later choice test. It remains, however, to be determined whether memories enhanced by large prediction errors may still bias decisions by prioritizing which experiences are sampled or reinstated during decision making.

In conclusion, we show that surprisingly large or small rewards and high-risk contexts improve memory, revealing that prediction errors and risk modulate episodic memory. We further demonstrated that absolute prediction errors have dissociable effects on learning rate and memory, pointing to separate influences on incremental learning and episodic memory processes.

2

Depressive Symptoms Bias the Prediction-Error Enhancement of Memory towards Negative Events in Reinforcement Learning[*]

In Chapter 1, we investigated the interaction between reward prediction errors and episodic memory and found that unsigned (absolute) prediction errors increase memory for a rewarding event, thereby prioritizing both more surprising positive and negative events in memory (Rouhani et al., 2018). It is unclear, however, how disorders marked by blunted positive and excessive negative affect, such as depression, may bias these effects on memory. To this end, we collected depression scores from all participants in our original sample, and tested for effects of depressive symptoms on reward learning, recognition memory, and the modulation of memory by prediction-errors.

Prior work characterizing reinforcement learning in Major Depressive Disorder (MDD) has demonstrated decreased sensitivity to rewards (Huys, Pizzagalli, Bogdan, & Dayan, 2013) as well as hypoactivation of reward-related responses in the striatum (for reviews, see Admon & Pizzagalli, 2015; Pizzagalli, 2014). Accordingly, attenuated reward prediction-error signals are reported in MDD (Gradin et al., 2011; Kumar et al., 2018); although these signals were intact in a task that did not require learning (Rutledge et al., 2017). Moreover, behavioral differences in reinforcement learning in MDD have been mixed. Many studies have shown similar learning performance between MDD patients and controls (Ubl et al., 2015) with differences modulated by individual levels of anhedonia (the inability to feel pleasure) independent of depression severity (Admon et al., 2015; Chase et al., 2010). In our non-

---

[*] The content of this chapter was published in Rouhani & Niv, 2019.

clinical sample, we therefore did not expect to see large differences in reward learning between those experiencing depressive symptoms and those that do not.

In addition to blunted reward processing, sustained negative affect in depression has led to work showing an asymmetry in processing negative over positive events. Namely, MDD patients show an attentional bias for negative stimuli, displaying difficulty in disengaging from and ignoring negative distractors (for reviews, see Gotlib & Joormann, 2010; Joormann & Quinn, 2014). In reinforcement learning, neuroimaging studies bolster evidence for this asymmetry by showing hyperactivation of cortico-striatal learning networks for punishment versus reward (Admon et al., 2015; Kumar et al., 2018; Ubl et al., 2015), including stronger prediction error signals for punishment (Kumar et al., 2018; Ubl et al., 2015). Of note, in depression, connectivity between the striatum and anterior cingulate cortex, a region associated with unsigned prediction errors (Roesch, Esber, Li, Daw, & Schoenbaum, 2012), is blunted in reward learning (Whitton et al., 2016), and enhanced in punishment learning (Admon et al., 2015).

These results suggest that in depressed individuals, high-magnitude negative prediction errors may have greater influence on learning and memory than do positive prediction errors. In line with this, depressed individuals exhibit a bias for negative versus positive memories (Gaddy & Ingram, 2014; Matt, Vázquez, & Campbell, 1992). This better memory for negative events in depression is thought to be modulated by the amygdala—a region associated with emotional memories as well as surprising events—and its functional connectivity with the hippocampus (Dillon, Dobbins, & Pizzagalli, 2014; Leal, Tighe, Jones, & Yassa, 2014; Sacchet et al., 2017; Young et al., 2017). Healthy individuals, on the other hand, exhibit a bias for positive versus negative memories, whereas depressed individuals additionally demonstrate an attenuated memory for positive events (Burt, Zembar, & Niederehe, 1995), which is linked to reduced activation in the dopaminergic midbrain and medial temporal lobes (Dillon et al., 2014). The literature therefore offers two mechanisms by which depressed individuals' memory may be biased compared to healthy individuals – better memory for negative events and worse memory for positive events. What remains to be elucidated is whether reward prediction error signals modulate this asymmetry in memory in depressed individuals.

To test this, we re-analyzed our previous results (see Chapter 1), and compared individuals reporting symptoms of depression (the "depressive" group) with those that reported no such symptoms ("non-depressive" group) on our measures of interest: their learning of the average values of the two scene categories, their trial-by-trial prediction errors and learning rates, and their recognition memory for those rewarding events.

2.1 METHOD

Across three experiments run on Amazon Mechanical Turk, 500 participants initiated the study (Exp 1: 200, Exp 2: 200, Exp 3: 100), 408 completed the study (Exp 1: 174, Exp 2: 148, Exp 3: 86), and after exclusions (see below), 383 participants are represented in our sample (Exp 1: 164, Exp 2: 136, Exp 3: 83). For details on the procedure, see Chapter 1. Importantly, at the end of each experiment, participants completed the Inventory of Depressive Symptomatology (IDS; Rush, Gullion, Basco, Jarrett, & Trivedi, 1996).

Statistical analysis

We investigated whether depression modulated learning and memory performance across all experiments. To do this, we compared participants who scored from moderate to very severe on the IDS (score: 26-84, which we refer to as "depressive", N = 101) to participants who reported low or no depressive symptoms (score: 0-13, which we refer to as "non-depressive", N = 184). Participants with an intermediate "mild" IDS score (14-25, N = 98) were excluded from the analysis (for categorization of scores, see www.ids-qids.org).

All comparisons were conducted using linear or generalized mixed-effects models (R lme4 package; Bates et al., 2015), with experiment as a random effect and subject as a nested random effect within experiment (for both intercept and slope), and trial-unique scene image as a random effect (for intercept). We used depression category (depressive or not) as a fixed or interacting effect to predict the below learning and memory measures. We additionally tested whether depression severity predicted the effects under question within the depressive group.

If depression was a significant predictor, to confirm group differences, we ran a simplified regression model (not including depression as an effect) separately within each experiment and depression group, and extracted subject-level intercepts and slopes. We then ran an ANOVA on the average difference in intercept and slope estimates between the "depressive" and "non-depressive" participants across all experiments. Finally, we corrected for multiple comparisons using Bonferroni correction.

Learning: As a measure of learning, we took the absolute deviation of participants' trial-by-trial estimates from the true average values of the scene categories (40¢ or 60¢ in Exp 1-2; 20¢ or 80¢ in Exp 3). This deviation should decrease as participants learn the average values of the scene categories. In other words, with every trial, the learner should be estimating closer to the true mean of that scene category, and so a significant effect of trial number in decreasing this measure reflects learning. We ran two models testing (1) whether depression predicted overall deviation from the true means and (2) whether depression interacted with trial number, indicating that depressed participants learned differently than non-depressed participants.

Prediction errors: Trial-by-trial prediction errors were calculated by subtracting participants' value estimates from the reward outcome experienced on that trial. We ran two models testing (1) whether depression predicted the average prediction error experienced during learning and (2) whether depression interacted with our previously reported finding that prediction errors are modulated by an interaction between risk context and scene-category value, leading to greater underestimation of the high-value category and greater overestimation of the low-value category in the high-risk room.

Learning rates: We calculated trial-by-trial learning rates as the proportion of the prediction error experienced on one trial that was then applied to update the value estimate on the next trial involving the same scene category. We ran four models testing (1) whether depression modulated the average learning rate applied during learning, (2) whether depression interacted with our previous findings that unsigned prediction errors increase learning rate, (3) whether depression interacted with our previous finding that a lower risk context leads to higher learning rates, and finally, (4) whether depression more specifically modulated an interaction between learning rate, unsigned prediction error and the valence of the prediction error (positive or negative) in a 3-way interaction; for example, surprising negative (versus positive) events could lead to higher learning rates (i.e., more value updating) in participants with depression.

Memory: We evaluated whether depression influenced item recognition by running the following mixed-effects logistic regressions predicting a "hit" or a "miss" during the memory test. We tested (1) whether depression affected overall memory, (2) whether depression interacted with the valence of the prediction error to influence memory; for example, by promoting negative prediction error memories over positive ones, (3) whether depression interacted with our previously reported finding that unsigned prediction errors increase memory, and (4) whether depression more specifically modulated an interaction between memory, the valence of the prediction error and absolute prediction error; for example, by selectively enhancing surprising negative events in memory over surprising positive ones.

2.2 RESULTS

Sample: Across all experiments, 184 participants scored within the "non-depressive" category (Exp 1: 69, Exp 2: 68, Exp 3: 47), and 101 participants scored within the "depressive" category (Exp 1: 51, Exp 2: 32, Exp 3: 18).

Learning: The absolute deviation of participants' estimates from the true averages of the two scene categories decreased as a function of trial number, indicating learning of the values of the two scene categories within each room (model 1: $\beta = -0.05$, $t = -2.93$, $p = 0.004$). Depression did not predict participant estimates on average (model 1: $\beta = -0.05$, $t = -1.04$, $p = 0.30$), nor did it interact with learning (model 2: $\beta = -0.02$, $t = -0.61$, $p = 0.54$); see Figure 2.1A-C. However, depression severity did predict an overall increase in estimation error within the depressive group (model 1: $\beta = 0.01$, $t = 3.44$, $p < 0.001$; Figure 2.1D).

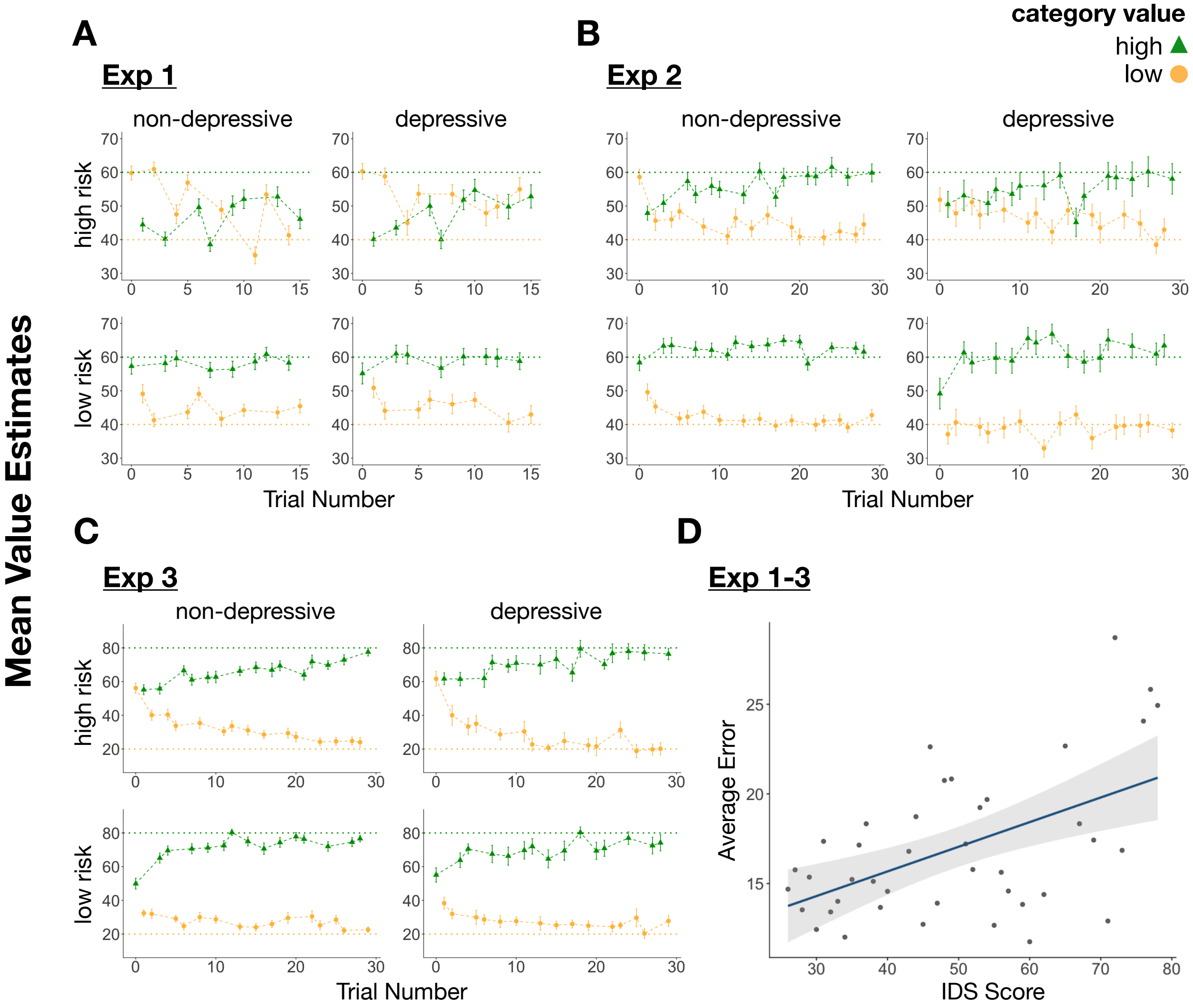


Figure 2.1: A-C: Learning. Average value estimates for high and low-value scene categories as a function of trial number, within high and low-risk rooms, divided between depressive and non-depressive groups, across all three experiments. We did not find any significant differences in value learning between the depressive and non-depressive groups. Note that "Trial 0" represents participant estimation at the beginning of each room and without having received any feedback. Error bars represent SEM. D. Average estimation error during learning as a function of IDS score in the depressive group. Depression severity predicted greater average error during learning. Each dot represents a participant, shaded regions represent 95% confidence intervals.

Prediction errors: Depression did not predict participant prediction errors (model 1: β = -0.00063, t = -0.033, p = 0.97), nor did it interact with the effect of risk and scene category value on prediction errors (model 2: β = -0.0062, t = -0.35, p = 0.73); see Figure 2.2.

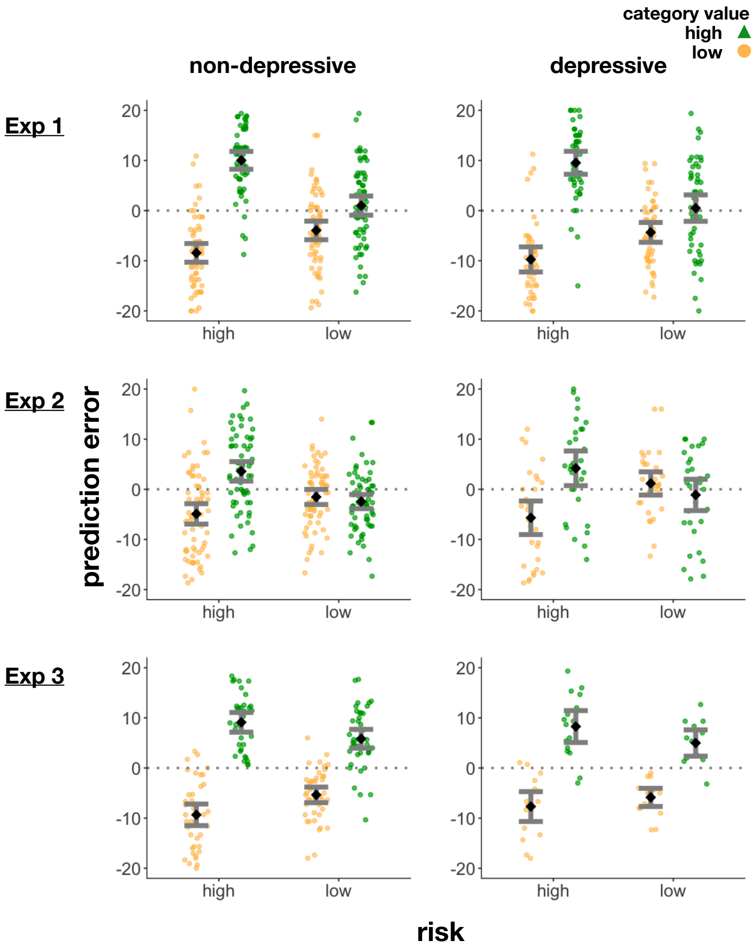


Figure 2.2: Prediction errors. Experienced prediction errors for high and low-value scene categories within high and low-risk rooms, divided between depressive and non-depressive groups, across all three experiments. There is an overall interaction between risk and category value, such that participants are more likely to overestimate the low-value category and underestimate the high-value category in the high-risk room. There were no differences between the depressive and non-depressive groups. Error bars represent SEM.

Learning rates: Trial-by-trial learning rates were similarly not predicted by depression (model 1: β = -0.046, t = -1.53, p = 0.13); depression did not interact with the increase in learning rate with unsigned prediction error (model 2: β = 0.034, t = 1.053, p = 0.29), nor did it interact with the effect of risk context on learning rate (model 3: β = -0.020, t = -0.92, p = 0.36). Finally, there was no effect of depression in an interaction between the unsigned value and the valence of the prediction error on learning rate (model 4: β = 0.034, t = 1.053, p = 0.29); see Figure 2.3.

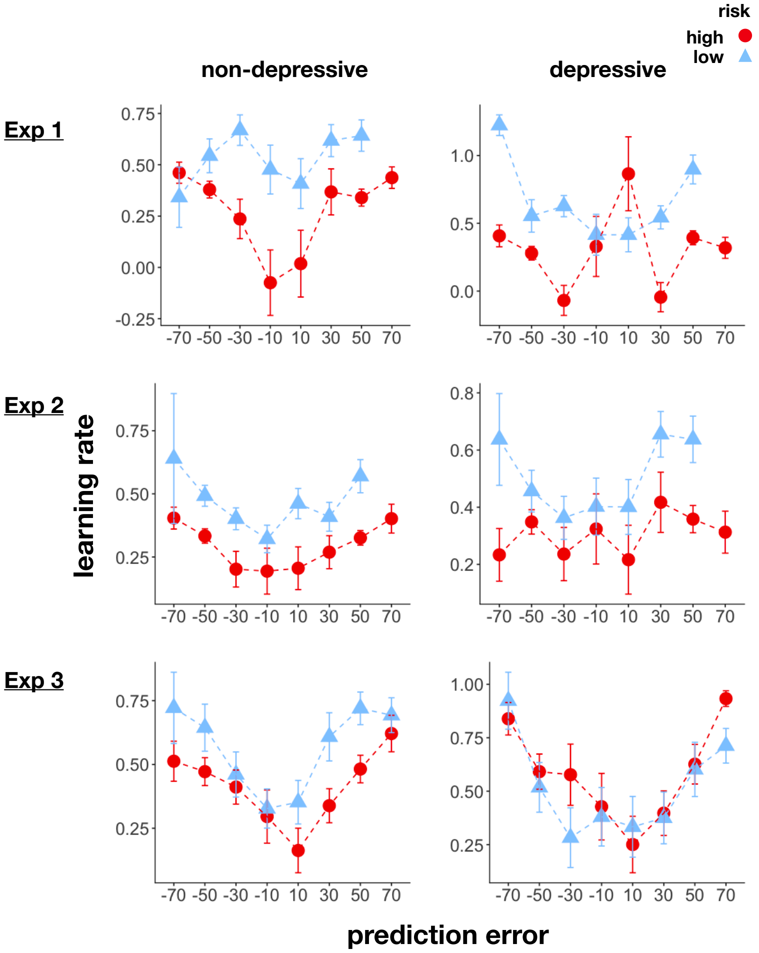


Figure 2.3: Learning rates. Average learning rates for outcomes as a function of prediction error in high and low-risk rooms, divided between depressive and non-depressive groups, across all three experiments. High magnitude prediction errors increase learning rate across all experiments and groups. We did not find any differences between the depressive and non-depressive groups. Error bars represent SEM.

Memory: Depression did not affect average recognition memory (model 1: $\beta = -0.019$, $z = -0.14$, $p = 0.89$). It did not interact with an effect of prediction error valence on memory in general (model 2: $\beta = 0.07$, $z = 0.53$, $p = 0.60$), nor with the effect of unsigned prediction error on memory (model 3: $\beta = 0.07$, $z = 1.02$, $p = 0.31$). However, we did find that depression modulated the interaction between the unsigned value and the valence of the prediction error on memory. In particular, "non-depressive" participants were more likely to remember more surprising positive events, while "depressive" participants were more likely to remember more surprising negative events, as predicted (model 4: $\beta = 0.31$, $z = 2.05$, $p = 0.040$; Figure 2.4).

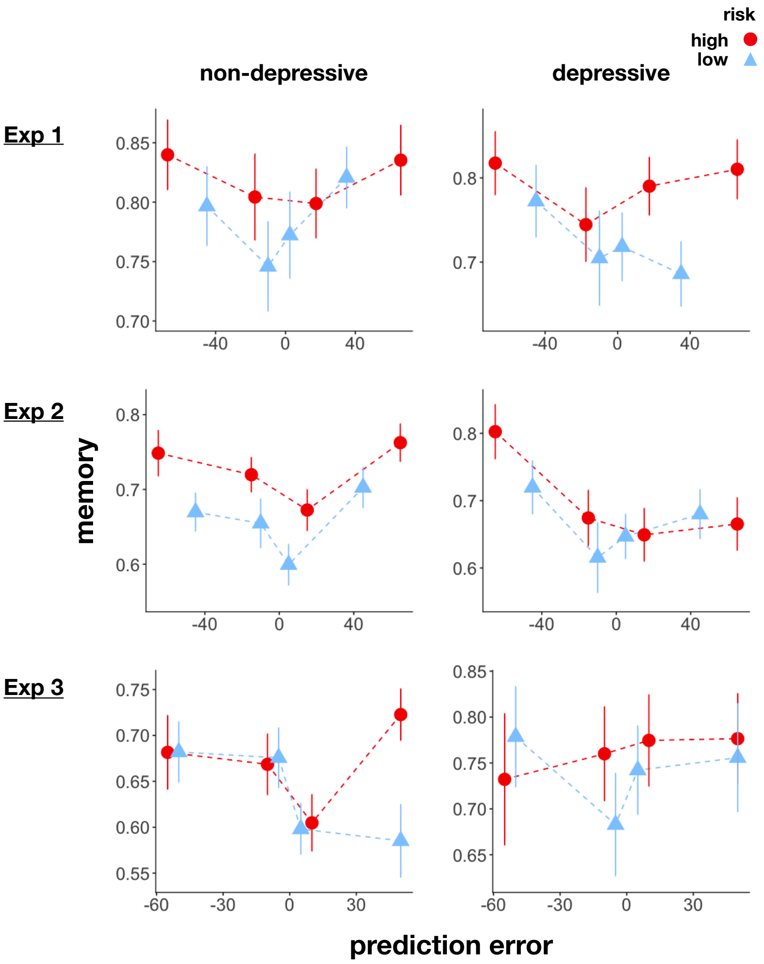


Figure 2.4: Memory. Binned item recognition memory as a function of prediction error in high and low-risk rooms, divided between depressive and non-depressive groups, across all three experiments. Item memory was binned by the quartile values of prediction errors within each room to illustrate the effects of prediction errors on memory; each dot represents the average value within that quartile. Note that no statistics were run on the binned values, and they are plotted only to illustrate the mixed-effects regression modeling. High magnitude prediction errors increased item recognition memory across all experiments and groups. There were no overall differences in memory between depression groups. However, there was a three-way interaction between the unsigned prediction error, the valence of the prediction error, and depression group, such that depressive participants are more likely to remember high-magnitude, negative prediction error items, whereas non-depressive participants are more likely to remember high-magnitude, positive prediction error items. Error bars represent SEM.

To confirm and further illustrate this effect (see "Statistical analyses" above for details), we found that an interaction between prediction error valence and depression group predicted the slope of the effect of unsigned prediction errors on memory ($F(1{,}283) = 16.95, p < 0.0001$). This interaction passed Bonferroni adjusted levels of $p = 0.004$ (alpha = 0.05/14 comparisons). Following up on this interaction, we tested for across and within-group differences. We found that depressive participants had higher slopes for negative prediction error events than non-depressive participants, $t(274.35) = 2.79$,

$p = 0.0057$, whereas non-depressive participants had higher slopes for positive prediction error events than depressive participants, $t(139.66) = -4.46$, $p < 0.001$. Within the depressive group, there were significantly higher slopes for the negative prediction error events than positive ones, $t(100) = -4.04$, $p < 0.001$, and within the non-depressive group, the opposite was true, $t(183) = 2.04$, $p = 0.043$. We did not find the interaction to predict the intercept of this model ($F(1,283) = 1.25$, $p = 0.26$); Figure 2.5.

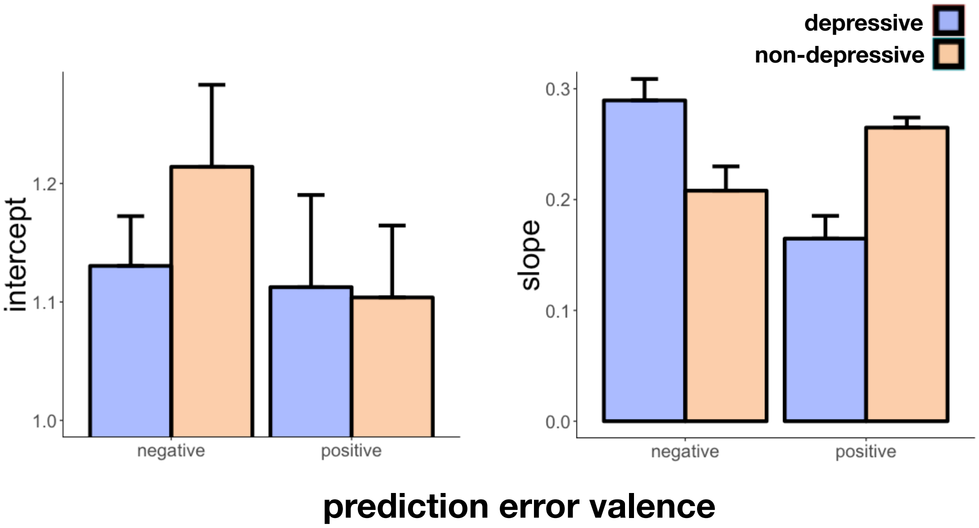


Figure 2.5: Intercept and slope values for the unsigned prediction error effect on memory. Mixed-effects logistic regression models were run separately for positive and negative prediction error outcomes in the depressive and non-depressive groups. Bar plots represent average intercept value (left) and slope value (right) as a function of the valence of the prediction error and depression group. There were no differences in the intercept value, but we found an interaction in the slope of this effect (representing the degree to which unsigned prediction errors are improving item memory), such that unsigned prediction errors are boosting memory more so for negative events in depressive individuals, and for positive events in non-depressive individuals. Error bars represent SEM.

2.3 Discussion

Depressive symptoms include a diminished ability to feel pleasure (anhedonia) as well as excessive negative affect, thereby suggesting abnormalities in learning from rewards as well as their effect on memory. In a non-clinical sample, we tested for differences in reward learning and memory performance in individuals with and without depressive symptoms. We did not find gross differences in learning performance between the two groups: subjects did not differ in how they learned the average values of two scene categories, as measured by their trial-by-trial estimates, prediction errors and learning rates throughout the task. However, we did find that depression severity predicted greater estimation error (i.e., difference between the estimated value and the true mean value of the scene categories) in the depressive group, which suggests impaired reinforcement learning in individuals with more severe depression. Nevertheless, we found that both groups increased their learning rates after high magnitude prediction errors, and neither group showed a bias towards updating expectations more after a positive or negative prediction error event. Together, these results suggest that dopaminergic prediction error signaling was relatively intact throughout our non-clinical sample.

We also did not find overall differences in memory performance, nor in memory for positive versus negative prediction error events, on average. Instead, we found that the modulation of memory by reward prediction errors was differently biased in the two groups such that in individuals with depressive symptoms, large negative prediction errors enhanced memory to a greater extent than did large positive prediction errors, and more so than they did in the non-depressive group. The opposite was true for non-depressive individuals: here, large positive prediction errors enhanced memories more than large negative prediction errors, and more so than they did in the depressive group.

Relatively intact learning in the depressive group is in line with several studies that have not found strong behavioral differences between MDD patients and healthy controls in reward learning (Chase et al., 2010; Knutson, Bhanji, Cooney, Atlas, & Gotlib, 2008; Smoski et al., 2009; Ubl et al., 2015). In our sample, however, the positive relationship between depression severity and task error suggests that reinforcement learning is affected in depression, but only in more severe cases, which may in part explain the heterogeneity of results in the literature. Moreover, we implemented a Pavlovian reward-learning paradigm that did not involve choices between differently rewarding options. This leaves open the possibility that depression is a greater modulator of instrumental learning than it is of prediction learning. Finally, given the striatal hypoactivity commonly reported in depression (for a review (Admon & Pizzagalli, 2015)), it is possible that depressive individuals are not as affectively influenced by reward, meaning they may not feel its associated pleasure or impact, even if they are unimpaired in following explicit task goals by using rewards to update the values of their experiences.

On the other hand, the surprise recognition memory test provides a measure unrelated to explicit task goals potentially capturing affect-driven cognitive biases in depression. Here, we did not find a general difference in memory for events associated with positive versus negative prediction errors between the depressive and non-depressive groups. We instead found a bias in the unsigned prediction error modulation of memory. This signal, which increases memory for surprising outcomes (Rouhani et al., 2018), more significantly modulated memory for negative prediction error events in the depressive group and positive prediction error events in the non-depressive group. In other words, depressive individuals were more likely to remember surprising negative events, whereas healthy individuals were more likely to remember positive ones. Such a bias in memory is in line with the tendency to ruminate on negative events in depression, and provides evidence that surprising negative (versus surprising positive) events are indeed prioritized in memory.

There are several mechanisms that could contribute to the better encoding of surprising negative events in depressive individuals. Unsigned prediction errors are known to increase arousal and deploy the LC-norepinephrine system (Nassar et al., 2012), which co-releases dopamine signals that induce hippocampal plasticity (Kempadoo et al., 2016; Takeuchi et al., 2016) and enhance episodic memory (Clewett et al., 2018). Our results therefore suggest that LC activity is modulated more by surprising negative events in depressive individuals and by surprising positive events in healthy individuals. Given projections between the LC and regions within the salience network, such as the anterior cingulate cortex and amygdala, previous work lends support to this hypothesis: depressive individuals show greater striatal-cingulate functional connectivity (Admon & Pizzagalli, 2015) and more amygdala-modulated memory for negative versus positive events, whereas the opposite pattern is true for healthy controls (Leal et al., 2014; Young, Siegle, Bodurka, & Drevets, 2016; Young et al., 2017).

Interestingly, in another line of work, the lateral habenula, which is associated with the processing of negative prediction errors (Matsumoto & Hikosaka, 2007), has been strongly implicated in modulating symptoms of depression (Yang et al., 2018). This link suggests that greater activity of the lateral habenula in depressive individuals may support the mnemonic bias towards negative prediction error events. Future neuroimaging work should characterize how unsigned prediction errors differentially modulate memory for negative versus positive prediction error events in depression. An alternative explanation is that an attentional bias for negative events (Gotlib & Joormann, 2010; Joormann &

Quinn, 2014) leads depressive participants to spend more time looking at scenes associated with strong negative prediction errors (and thereby encoding them in memory). Future studies could test this by using eye-tracking as a measure of attention.

Our study has several limitations. First, our depressive group was not a clinical sample, and our findings need to be tested specifically in patients suffering from MDD. Moreover, given the heterogeneity of symptoms in MDD, future studies should take additional measures to allow testing for the modulation of the interaction between learning and memory by the severity of symptoms such as anhedonia, rumination, and anxiety. In particular, anhedonia has been shown to impair reward learning performance regardless of depression severity (Admon & Pizzagalli, 2015; Chase et al., 2010; Gradin et al., 2011), and can similarly desensitize individuals to negative outcomes, whereas anxiety increases sensitivity to negative outcomes (Mueller, Pechtel, Cohen, Douglas, & Pizzagalli, 2015). Individual measures of depressive symptoms along with co-morbid symptoms of anxiety could provide a better picture of which aspects of the disorder are giving rise to the biases in memory. We additionally did not collect medication information so could not test for the potential effects of neuroactive substances on learning and memory performance.

Nevertheless, it is notable that 26% of our Amazon Mechanical Turk (mTurk) sample scored moderately to severely depressed. This is in line with a recent finding that depression is two to three times higher in mTurk workers (under 50 years old) than matched national samples (Walters, Christakis, & Wright, 2018). This further suggests that researchers can characterize or, alternatively, need to control for the effects of depression in their mTurk experiments.

In conclusion, our findings, in a heterogenous, online, non-clinical population are fully in line with previous literature, suggesting that our task, and the interactions we found between learning and memory, may prove useful in clinical settings as well.

3

Two Reward Prediction Error Signals Dynamically Modulate Memory during Distinct Phases of Reinforcement Learning

Here, we turn to classic associative models of learning to characterize an attentional mechanism supporting the reward prediction error (RPE) enhancement of learning and memory reported in Chapter 1. These associative models of learning explain how attention to cues change as a function of experience (Pearce & Mackintosh, 2010), but it is less clear how such a mechanism may influence memory. We draw on two well-known associative models of attention, the Mackintosh model (Mackintosh, 1975) and the Pearce-Hall model (Pearce & Hall, 1980), and test whether they predict learning and subsequent memory for events experienced during reinforcement learning. Quizzically, these two models make opposite predictions for behavior. In the Mackintosh model, attention increases for cues with strong associative strengths (i.e., cues that reliably predict reward), whereas in the Pearce-Hall model, attention is enhanced for cues that are accompanied by surprise, that is, those that co-occur with large unsigned prediction errors (i.e., cues that unreliably predict reward). Albeit paradoxical, both theories are justified by decades of evidence and are thought to represent two distinct attentional mechanisms differentially deployed given the demands of learning (for a review, see Pearce & Mackintosh, 2010). We treat signed and unsigned RPEs as drivers of these two attentional signals, and dissociate their effects on memory during different phases of reward learning. We thus provide an answer to the open question (Ergo, De Loof, & Verguts, 2020) of whether and how both signed and unsigned RPEs modulate memory during reinforcement learning.

In these models, a dynamic learning rate, termed "associability", represents attentional fluctuations during learning which, in turn, govern the value updating of conditioned stimuli. In the Pearce-Hall model, associability increases in proportion to unsigned prediction errors (i.e., the degree of surprise), so that unreliable cues are attended to while learning resources are not directed to cues that reliably predict outcomes (Pearce & Hall, 1980). In the Mackintosh model, on the other hand, the stronger the predictive strength, the higher the associability and attention to the cue (Mackintosh, 1975). It is theoretically possible that greater cue-outcome 'surprise' and predictability both modulate attention throughout learning (Le Pelley, 2004; Beesley, Nguyen, Pearson, & Le Pelley, 2015).

We used reinforcement learning models to test for the influence of each of these attentional components on dynamic (trial-by-trial) learning rates. We treated the unsigned RPE at reward outcome as a 'Pearce-Hall' signal, as it reflects how unpredictable the reward was. We treated the learned value difference between two reward-predictive cues as a 'Mackintosh' signal, where greater predictiveness is equated with higher learned values for one cue versus the other. We refer to this value signal as a signed RPE experienced at cue, as when there are several possible cues, the onset of a cue resolves the prediction for the current trial, and is accompanied by an RPE that reflects the difference between the current predicted reward, and the average reward predicted before cue onset (Niv & Schoenbaum, 2008).

We were furthermore interested in how these RPEs, one evoked by cue and the other by reward outcome, influence memory for events experienced at cue and outcome. In our previous study (Chapter 1; Rouhani et al., 2018), we found that unsigned RPEs at outcome boost memory, in line with Pearce-Hall and putatively signaled by noradrenaline from the locus coeruleus (LC, co-releasing dopamine; Takeuchi et al., 2016; Kempadoo et al., 2016; Wagatsuma et al., 2017). We did not find signed RPEs at outcome, canonically associated with phasic dopaminergic signals (Barto, 1995; Montague et al., 1996), to influence memory.

Nevertheless, there is evidence supporting signed-RPE effects on memory during reinforcement learning (Davidow et al., 2016), including a recent study where the higher (i.e., more positive) the RPE, the greater the memory enhancement (Jang et al., 2019). Interestingly, however, this positive RPE occurred during the reward-predictive cue, at the point when participants realized their likelihood of receiving reward but prior to experiencing the actual outcome, likening this effect to work showing memory benefits during periods of high-reward anticipation (Adcock et al., 2006; Murty & Adcock, 2014; Stanek, Dickerson, Chiew, Clement, & Adcock, 2019; Wittmann et al., 2005). Of course, high reward expectations depend on having learned the values of the predictive cues. Accordingly, this memory enhancement is thought to be initiated by the signed RPE signal, which transfers from reward outcome to cue with more learning (Schultz, Dayan, & Montague, 1997), supporting the Mackintosh signal described above. Our previous study included the same event at cue and outcome, potentially obscuring signed-RPE effects at cue (Rouhani et al., 2018).

In the following experiments, we therefore included two trial-unique events on every learning trial, one at reward cue and one at outcome, to dissociate the effects of two RPEs, one experienced at reward cue and one at outcome, on memory for these distinct reward-learning events (Figure 3.1).

Overview of experiments

We characterized the effects of a cue and outcome RPE on learning and memory in two experiments that each prioritized the influence of one of these RPE signals. In Experiment 1, participants learned the value of a single reward category and experienced large unsigned RPEs brought on by periods of high outcome variance (‘high’ versus ‘low variance’ contexts) and reward value change-points (changes to the mean of the underlying reward distribution). We expected these large unsigned RPEs, experienced at reward outcome, to modulate learning rate (as in Pearce-Hall models; Pearce & Hall, 1980) and to boost memory for events throughout learning.

In Experiment 2, on the other hand, participants learned the values of two reward categories, eliciting RPEs at cue as well as at reward outcome. Here, the underlying reward distribution associated with each category did not change, allowing for RPEs at cue (i.e., a relative value signal) to increase in magnitude with more experience with each reward category. We moreover included two learning conditions that varied in learning difficulty due to different degrees of overlap between the reward distributions of the two categories. In Condition 1, the means of the two reward categories were close together (40¢ and 60¢) with considerable overlap in their reward distributions. In Condition 2, on the other hand, the two means were further apart (20¢ and 80¢) and there was no overlap between the two reward distributions. We predicted that cue RPEs would increase in size as a marker of learning in both conditions; however, in Condition 2, the greater separation of the two reward distributions should lead participants to experience larger and earlier cue RPEs than in Condition 1. Regardless of condition, we expected signed RPEs at cue to influence learning rate (as in the Mackintosh model; Mackintosh, 1975), in addition to unsigned RPEs at outcome (as in the Pearce-Hall model; Pearce & Hall, 1980). We also expected this signed cue RPE to boost memory for the more valued reward category with more learning, reflecting the mnemonic effects of stronger reward expectations and anticipation.

In sum, we hypothesized that an unsigned RPE at outcome (as in a 'Pearce-Hall' signal of attention) as well as a signed RPE at cue (as in a 'Mackintosh' signal of attention) would enhance learning and memory for both more surprising and more valued events (Figure 3.1). We additionally characterized the effects of a signed RPE at outcome – the learning signal putatively giving rise to the signed RPE we test at reward cue – as well as an unsigned RPE at cue, on memory. In fact, in our paradigm, an unsigned RPE at cue could also be thought of as a Mackintosh signal, as larger RPE magnitudes at cue indicate greater learned separation between the values of two reward categories, reflecting stronger associative strengths for each cue. Given our previous results (Rouhani et al., 2018), we did not expect a signed RPE at outcome to influence memory. However, we did expect that as participants learn to separate the values of different reward categories, represented by larger unsigned RPEs at cue, this increased predictiveness ('Mackintosh' signal) would lead to better memory for those cue events.

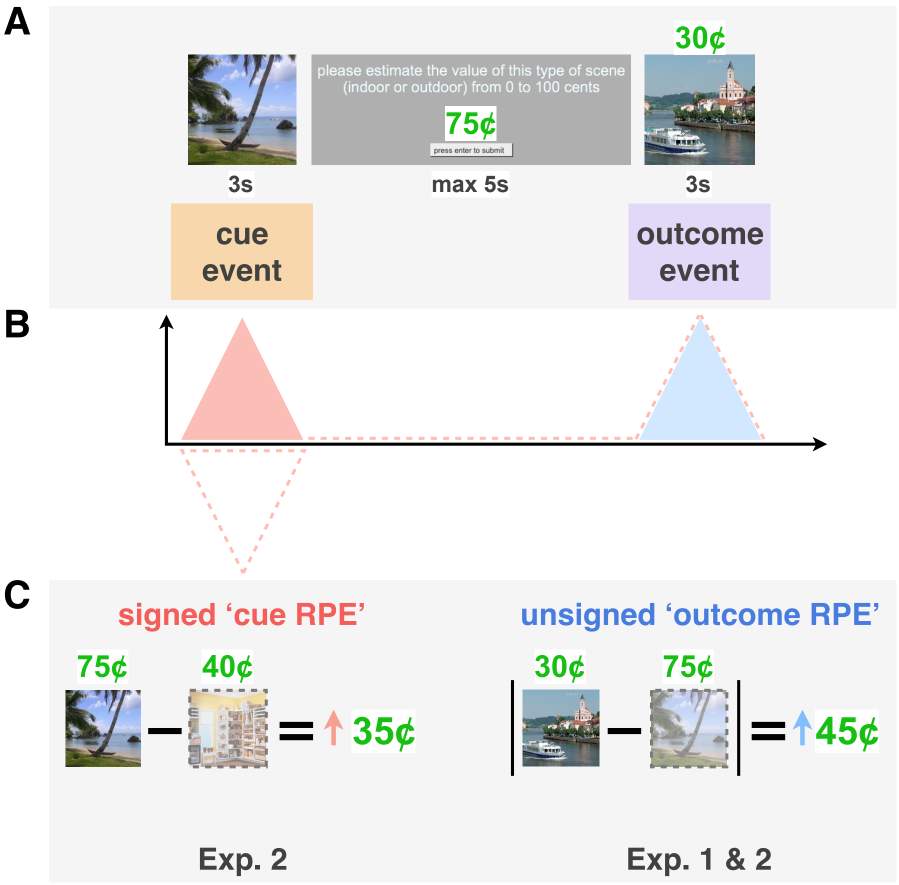


Figure 3.1: Reward prediction error (RPE) signals in a learning trial. **A**. Each trial was initiated by a reward cue represented by a trial-unique event, participants were then asked to indicate how much ‘on average’ that reward category was worth, after which they saw the reward outcome (a proportion of which they received) along with a different trial-unique event. In Experiment 1, all cue and outcome events were objects (single reward category), whereas in Experiment 2 (as pictured), each trial included either two indoor or two outdoor scene events (two reward categories). **B-C**. Theoretical RPE signals during a learning trial (B) and their calculation (C). During learning, the signed, dopaminergic, RPE signal moves from reward outcome to the cue predicting the outcome (phasic response in red, journey from outcome to cue in dotted line). We calculated this signal at cue by subtracting the value of the current reward category (outdoor scenes) from the other category (indoor scenes); note that this value can be negative (dotted triangle). We expected this signed signal (in red) to boost memory for more valued events. We tested the effects of this cue RPE in Experiment 2, as a cue RPE requires learning of more than one reward category. We also calculated an unsigned, putatively noradrenergic, RPE at outcome (in blue) by taking the difference between the value for that reward cue and its subsequent outcome. We expected this unsigned signal (in blue) to enhance memory for more surprising outcomes, which we tested in both Experiments 1 and 2.

3.1 Reinforcement learning and memory models

To determine how unsigned RPEs at reward outcome (‘outcome RPE’) and signed RPEs at reward cue (‘cue RPE’) influence learning, we modeled participants’ trial-by-trial value estimates testing a series of reinforcement learning models. We used a simple Rescorla-Wagner model (Rescorla & Wagner, 1972) as our baseline model (model: ‘RW’):

$$V_{t+1} = V_t + \alpha(R_t - V_t), \tag{3.1}$$

where a static learning rate (α) governs the extent to which the signed RPE at outcome (computed by subtracting the current model value, V_t, from the reward received on that trial, R_t) updates the value of the next trial (V_{t+1}).

Following attentional models of learning (Pearce & Mackintosh, 2010), we investigated whether a dynamic trial-specific learning rate (α_t) would better fit learning. We tested three distinct modulators of a trial-by-trial learning rate, separately and in combination with each other. To constrain α_t to be in the range of [0-1], for each model, we passed the learning rate through a sigmoid function before updating value (Eq. 3.1).

First, in line with Pearce & Hall (Pearce & Hall, 1980), we used the unsigned (absolute) outcome RPE to modulate learning rate (model: 'RW-PH'):

$$\alpha_t = \eta + \kappa(|R_t - VE_t|). \tag{3.2}$$

Here, the unsigned outcome RPE is calculated as the difference between the reward received and the participant value estimate (VE_t). The learning rate is set as a baseline learning rate, η, plus the unsigned RPE scaled by κ. For positive values of κ, more surprising outcomes therefore lead to higher learning rates, as per the Pearce-Hall model.

Second, following Mackintosh (Mackintosh, 1975), we modeled the effect of a cue RPE on learning rate (model: 'RW-M'). Note that we could only test this effect in Experiment 2 since cue RPEs exist only when there is more than one reward category. The cue RPE is the value of the present reward category (e.g., an indoor scene; VE_c) relative to the value of the other reward category (e.g., an outdoor scene; VE_n). The learning rate in this model is then the scaled cue RPE plus a baseline learning rate η:

$$\alpha_t = \eta + \gamma(VE_c - VE_n), \tag{3.3}$$

Therefore, for positive γ, the more one scene category is valued over the other, the higher α_t for trials with the more valued scene category and the lower α_t for trials with the less valued scene category[*].

Third, given that participants should update their values less (i.e., lower their α_t) once they've learned the average values of the reward categories, we tested a model with exponential decay of the learning rate over time (Sutton & Barto, 1998; model: 'RW-D'):

$$\alpha_t = \eta + Ne^{-\lambda t_c}, \tag{3.4}$$

where N is the initial value, λ is the decay constant, and t_c is the trial number for that reward category (i.e., in Experiment 2 where there were two scene categories, trial number was counted separately for each scene category).

We further tested models that included each combination of the above three learning-rate modulators. Here, we used a single baseline (η) and added each effect in the learning rate for all of the following models: A model that combines the unsigned outcome RPE and signed cue RPE effects on learning rate (model: 'RW-PH-M'):

$$\alpha_t = \eta + \kappa(|R_t - VE_t|) + \gamma(VE_c - VE_n), \tag{3.5}$$

A model that combines the unsigned outcome RPE and decay effects on learning rate (model: 'RW-PH-D'):

$$\alpha_t = \eta + \kappa(|R_t - VE_t|) + Ne^{-\lambda t_c}, \tag{3.6}$$

A model that combines the signed cue RPE and decay effects on learning rate (model: 'RW-M-D'):

$$\alpha_t = \eta + \gamma(VE_c - VE_n) + Ne^{-\lambda t_c}, \tag{3.7}$$

And finally, a model that combines all three effects (model: 'RW-PH-M-D'):

$$\alpha_t = \eta + \kappa(|R_t - VE_t|) + \gamma(VE_c - VE_n) + Ne^{-\lambda t_c}. \tag{3.8}$$

---

[*] Since each scene category was sampled an equal number of times (without any runs exceeding two trials), we did not scale the cue RPE by the probability of either scene category occurring.

Model fitting and comparison

All models were fit to each participant's value estimates by finding parameters that maximize the log likelihood of the participant value estimates. The likelihood was calculated assuming a Gaussian distribution around the model value, with variance equal to the empirical variance between model values and participant estimates (σ^2). This is equivalent to linear regression of the value estimates on the model values, giving a log likelihood:

$$LL = -n_{data}\left[ln\left(\sqrt{2\pi\sigma^2}\right) + 0.5\right], \tag{3.9}$$

where n is the number of trials fit. To maximize log likelihood we used MATLAB's *fmincon* function. We constrained parameter values within the following ranges: $\alpha \in$ [0,1], $\eta \in$ [-10,10], $\kappa \in$ [-20,20], $\gamma \in$ [-20,20], $N \in$ [-15,15], $\lambda \in$ [-20,20]. Note, however, that the trial-by-trial learning rate was always passed through a sigmoid function, and was therefore between 0 and 1. Values were initialized to 50¢, and in Experiment 1, were re-initialized at the beginning of each reward context. Each fit was run 30 times with different random initial parameter values.

Since all our models were nested (with additional parameters further modulating the RW-learning rate), we compared models using the likelihood-ratio test (Pickles, 1985). We also compared models using the more conservative Bayesian information criterion (BIC; Schwarz, 1978). Both tests were conducted across subjects.

HIERARCHICAL MODEL OF MEMORY

We ran a hierarchical regression model to better characterize the effects of unsigned and signed RPES, as well as their relative influence, on memory for cue and outcome events. This model performed full Bayesian inference over the effects of interest with Hamiltonian Monte Carlo sampling, simultaneously estimating subject and group-level posterior distributions (Stan; Carpenter et al., 2017). We included all putative RPE signals of interest in predicting memory score: signed RPE signal at outcome, unsigned RPE signal at outcome, as well as an intercept and a nuisance variable that captured overall differences in memory for cue versus outcome events. We also included signed and unsigned

RPE signals at cue for Experiment 2. Subject-level parameter distributions were drawn from group-level, standard normal distributions, and scaled by a gamma distribution (1,0.5). The response variable (memory score) was modeled with a normal distribution and fit with a single Gaussian noise parameter across all participants. All RPE regressors were centered and standardized. We report the median (M) of the posterior parameter distributions as a measure of centrality, and the highest density interval (HDI) as a measure of uncertainty around the parameter estimate; by default, HDI returns the 89% credible interval (which is further recommended as a more stable interval for sample sizes less than 10,000; Kruschke, 2014; Makowski, Ben-Shachar, & Lüdecke, 2019).

3.2 EXPERIMENT 1

Method

Participants

One-hundred participants were recruited from Amazon Mechanical Turk (MTurk). The sample size was chosen (1) based on a simulation-based power analysis revealing that at least 55 participants would give sufficient power (80% probability) to detect the effect of unsigned RPEs on memory (Rouhani et al., 2018), and (2) taking into account that 20% of participants typically meet one of the following exclusion criteria. Participants were excluded if they (1) had a memory score of less than 0.5 (A': Sensitivity index in signal detection; Pollack & Norman, 1964), or (2) missed more than three trials. This led to a final sample of 81 participants. We obtained informed consent online, and participants had to correctly answer questions checking for their understanding of the instructions before proceeding; procedures were approved by Princeton University's Institutional Review Board.

Task design

Participants each completed (1) a reward learning task, (2) a recognition-memory task, and (3) a choice task. They were additionally asked to complete a risk questionnaire (DOSPERT; Weber et al., 2002) between reward learning and memory to create a 5-10 minute delay between item encoding and recognition.

Reward learning: Participants learned the average value of objects in two different reward contexts, defined by background images of different cities ('Paris' and 'London'). They experienced each reward context in interleaved blocks (8 blocks total). Each block was comprised of 6 or 9 trials (60 trials total), each trial involving two trial-unique objects (120 objects in total). On each trial, participants were first shown an object ('reward cue': 3 seconds), and then had up to 5 seconds to estimate the "resale value of objects in that city at that time", i.e., the average value of objects in that context. After submitting their answer, they saw a different trial-unique object ('reward outcome': 3 seconds) along with the reward value associated with both objects on that trial. Participants were paid 10% of the rewards they received on every trial regardless of their estimates, in line with a Pavlovian conditioning environment.

The individual rewards associated with the object pairs fluctuated around a fixed mean (the means ranged from 10¢ to 90¢). Once or twice within each reward block, the underlying mean changed, generating large RPEs. These 'change points' occurred once in the 6-trial blocks, twice in the 9 trial blocks, and were at least 3 trials apart. The reward variance associated with each context provided a second source of RPEs. The variance in the high-variance context (σ-high-variance = 7¢) was twice that of the low-variance context (σ-low-variance = 3.5¢), leading participants to experience larger RPEs within the high-variance context. Participants were told that the average resale value of the "found" objects could change within each city, but that the inherent variability in reward outcome associated with each city remained constant. Participants were encouraged to remember the rewards associated with the objects, as they were told they would be choosing between objects, and re-earning their associated rewards, later in the task.

Recognition memory: After completing the risk questionnaire, participants were tested for their memory of the trial-unique objects. They were presented with objects and asked to indicate whether they were 'old' or 'new' as well as their confidence level for each memory judgment (from 1 'guessing' to 4 'completely certain'). The test included 72 trials: 48 old (24 from each context) and 24 new images.

Choice: In the final phase, participants were asked to choose the more valuable object between two previously seen objects (14 trials total). Unbeknownst to the participants, objects within each pair were either (1) close in their associated reward but belonged to different variance contexts (6 trials), or (2) belonged to the same pair and were therefore associated with the exact same reward (8 trials). Any consistent biases in preference between reward events (as represented by the objects) could therefore not be attributable to explicit reward differences in the task. Choice results are reported in Section 3.4.

Statistical analysis

We analyzed learning, memory and choice data using complementary approaches. We tested putative effects of signed cue RPEs and unsigned outcome RPEs on learning by generating and comparing computational models of learning (specifically modeling trial-by-trial value estimates). We tested these effects on memory performance through mixed-effects modeling (also used to analyze choices lme4 package; Bates et al., 2015) and Bayesian hierarchical modeling in Stan (Carpenter et al., 2017). Trial-by-trial memory scores were calculated by combining memory performance (hit versus miss) with confidence rating (from 1 = 'guessing' to 4 = 'completely certain') on old items; the score thus ranged from a 'completely certain' miss (1) to a 'completely certain' hit (8).

Reinforcement learning results

We fit learning behavior using four models: 'RW', 'RW-PH', 'RW-D', and 'RW-PH-D'. Note, we did not test any model that included a cue RPE since there is a single reward category in this experiment. We found that a model that included a Pearce-Hall component, and therefore modulated learning rate by the unsigned RPE at outcome, fit better than a model without that component (paired t-test on BIC score differences for RW-PH versus RW, $t(80) = 3.43, p < 0.001$; Table 3.1, Figure 3.2B). The BIC score was marginally better for the less complex RW-PH model relative to the RW-PH-D model ($t = 1.76, p = 0.08$). However, the RW-PH-D model fit the data significantly better than the RW-PH model as per the likelihood ratio test ($\chi^2(162) = 212.87, p = 0.005$). We therefore cannot adjudicate between these two models; however, we can conclude that a model including the modulation of learning rate by unsigned RPE at reward outcome, in line with Pearce and Hall (Pearce & Hall, 1980), better explained learning behavior in our task.

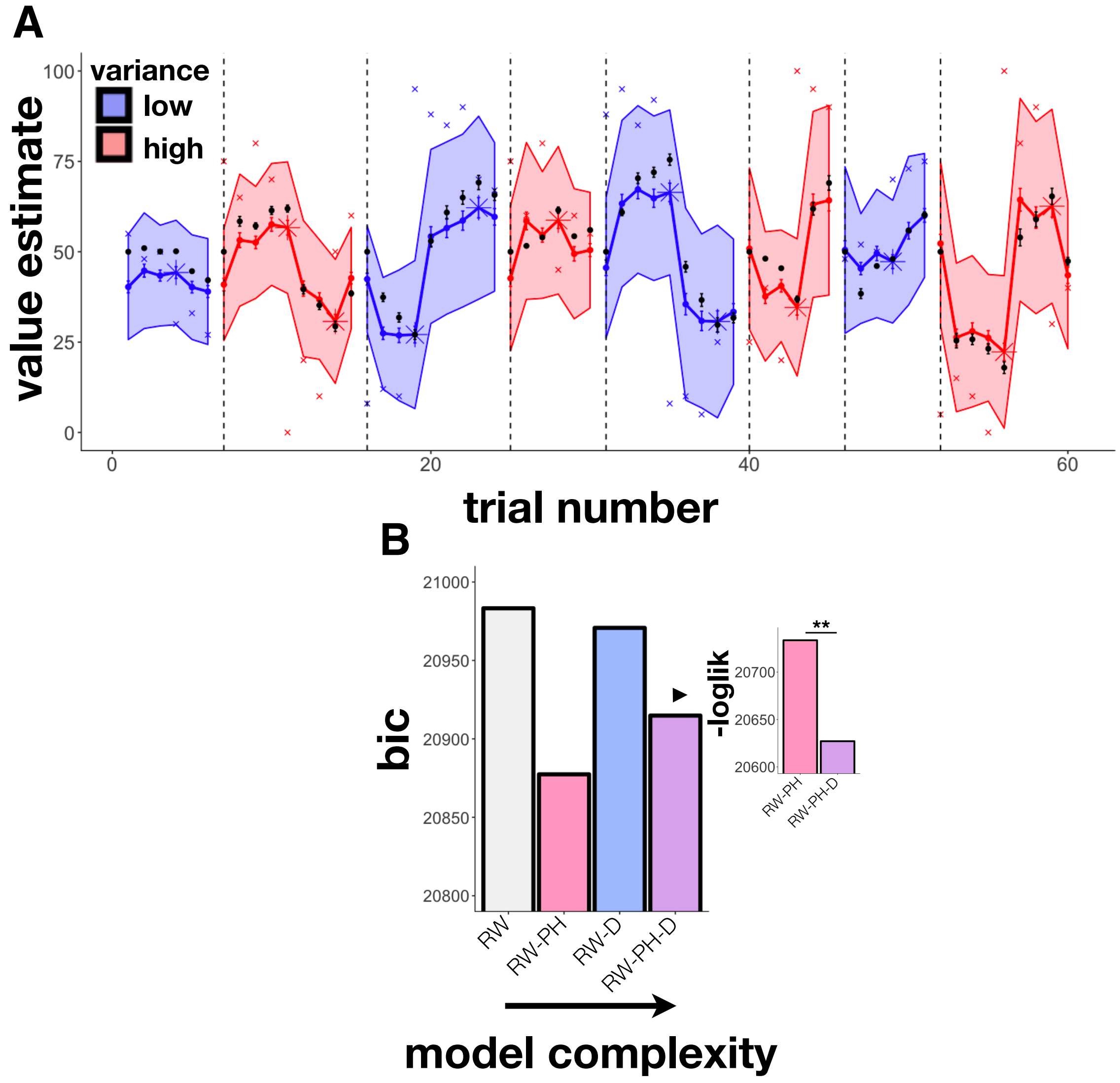


Figure 3.2: Experiment 1 learning behavior and modeling results. **A**. Experiment 1 average participant value estimates as a function of trial number (red and blue lines; stars indicate a change-point trial; shading indicates 95% confidence intervals), and average model value predictions of the RW-PH-D model with SEM bars in black. True rewards on each trial are indicated by x's. **B**. Total BIC scores across subjects for models tested in Experiment 1. Lower scores indicate better fit; insets show negative log-likelihoods for the two best fitting models, as tested in the likelihood-ratio test. In Experiment 1, we cannot adjudicate between the two best models, as BIC scores are better for the RW-PH model but the likelihood-ratio test prefers the RW-PH-D model; nevertheless, we can conclude that models that included a PH component outperformed models that did not.

Memory results

Memory by learning condition

We first tested the effects of our task conditions on memory, namely how reward variance modulated memory for cue and outcome events. In line with our previous study (Rouhani et al., 2018), we expected that the larger unsigned RPEs in a high-variance context would improve memory for especially those outcome events. Moreover, given that we only included a single reward category in this experiment and therefore did not elicit RPEs at cue, we did not expect memory for cue to substantially change through learning. We, however, did suspect overall better memory for cue relative to outcome events, since the additional monetary outcome that accompanies the outcome event may, in general, interfere with its encoding; we control for this nuisance effect in all of our analyses.

Accordingly, we found consistent better memory for cue versus outcome events throughout learning (mixed effects linear regression: β = -0.84, t = -10.23, p < 0.001; Figure 3.3A). We also found an interaction of cue versus outcome memory by variance condition, such that in the high-variance condition, there was a lower average memory score for cue events, and a higher average memory score for outcome events, compared to the low-variance condition (μ-high-cue = 6.44, μ-low-cue = 6.57, μ-high-outcome = 5.79, μ-low-outcome = 5.54; β = -0.37, t = -2.78, p = 0.005). Within the interaction, there was a significant difference in memory for outcome events (β = -0.25, t = -2.09, p = 0.04) but not for cue events (β = 0.12, t = 1.41, p = 0.16). This suggests a role for the high-variance context, characterized by larger unsigned RPEs, in boosting memory for outcome events. See below, 'Memory by RPE signals', for a direct test of the effect of RPEs on memory for outcome events.

Memory by RPE signals

Replicating our previous findings (Rouhani et al., 2018, 2020), we found that unsigned, but not signed, outcome RPEs increased memory for associated events (mixed-effects logistic regression, |outcome RPE|: β = 0.15, t = 4.13, p < 0.001; outcome RPE: β = -0.03, t = -0.76, p = 0.45; model controlled for differences in memory for cue and outcome events). This effect was significant in both variance conditions, and larger in the high-variance condition, which was characterized by larger outcome RPEs (high variance: β = 0.20, t = 3.85, p < 0.001; low variance: β = 0.10, t = 2.01, p = 0.04).

When using the hierarchical model to characterize the effects of RPE signals on memory for cue and outcome events in Experiment 1, we found that unsigned outcome RPEs enhanced memory for both events, but boosted memory for outcome events more so with a median parameter estimate al-most twice the size of that for cue memory (outcome memory as a function of |outcome RPE|: M = 0.14, HDI [0.06, 0.23], $\beta = 0.19$, $t = 3.51$, $p < 0.001$; cue memory as a function of |outcome RPE|: M = 0.08, HDI [-0.01, 0.16], $\beta = 0.10$, $t = 2.29$, $p = 0.02$; Figure 3.3B). As expected, we did not find that signed outcome RPE influenced memory for cue or outcome events (outcome memory as a function of outcome RPE: M = -0.0007, HDI [-0.09, 0.08], $\beta = -0.004$, $t = -0.07$, $p = 0.95$, Figure 3.3C; cue memory as a function of outcome RPE: M = -0.02, HDI [-0.11, 0.06], $\beta = -0.05$, $t = -1.04$, $p = 0.30$).

Finally, since unsigned outcome RPEs in this experiment were modulated by both reward variance as well as change-points to the underlying distribution of the mean, we checked for the individual effects of reward variance and change-points on outcome memory. We found separate effects of each in increasing memory for outcome items (outcome memory as a function of |outcome-RPE|: $\beta = 0.14$, $t = 2.38$, $p = 0.02$; outcome memory as a function of change-point: $\beta = 0.24$, $t = 2.08$, $p = 0.04$), showing that this effect cannot be fully explained by change-point events, and instead, also relies on the magnitude of the outcome RPE.

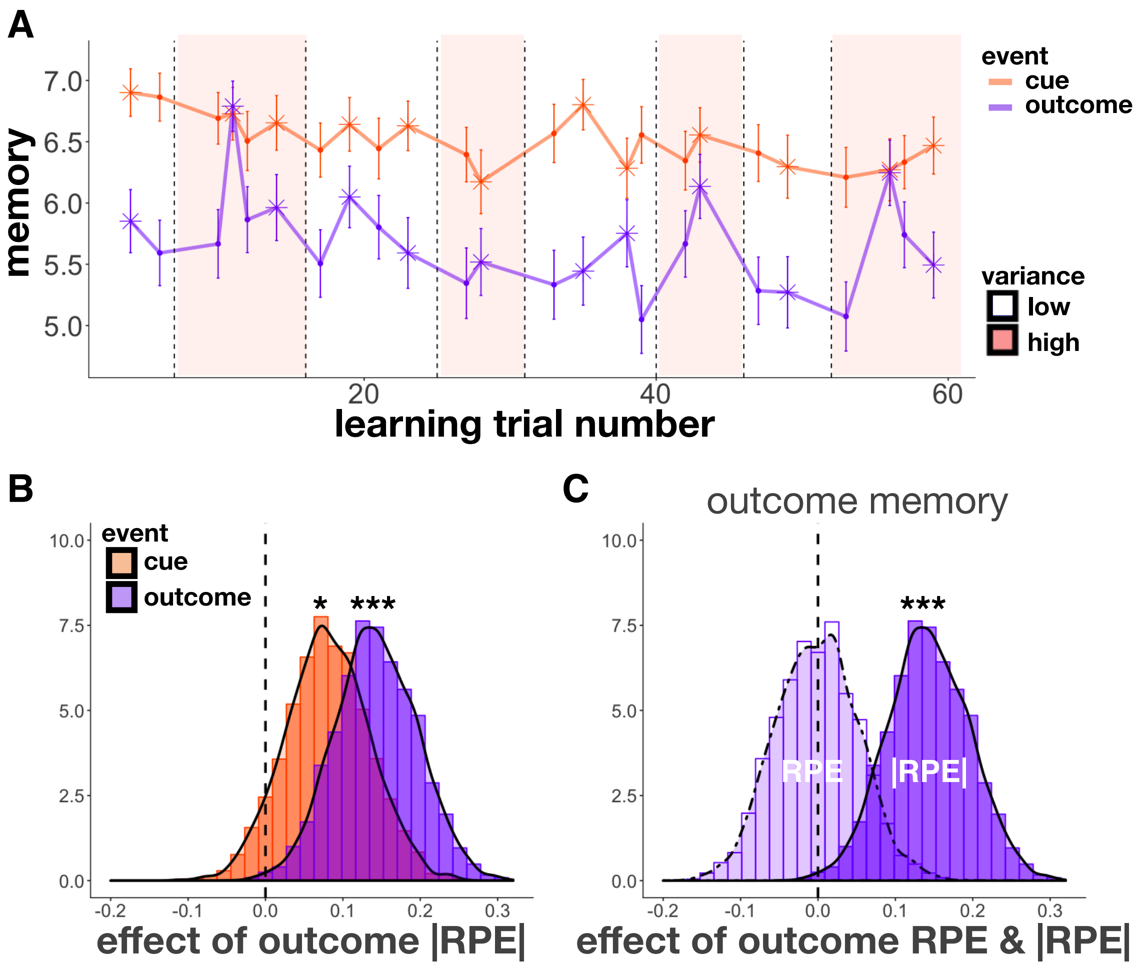


Figure 3.3: Experiment 1 memory results. **A**. Experiment 1 memory score as a function of trial number; starred points indicate change-point events. Background shading indicates condition (low or high variance). Cue memory was in general better than outcome memory, with an interaction between this difference and variance condition. **B-C**. Parameter distributions from hierarchical models of memory. Experiment 1: (B) Unsigned outcome RPEs effect on cue and outcome memory. Unsigned outcome RPEs enhanced both cue and outcome memory, and this effect was stronger for outcome memory; (C) RPE effects on outcome memory. Signed outcome RPEs did not influence memory (unsigned RPE plot identical to panel B, repeated for comparison with signed RPE).

Memory by learning rate

In Chapter 1 (Rouhani et al., 2018), we found that although unsigned RPEs at outcome enhanced both learning rate and memory, learning rate did not predict subsequent memory for those events, suggesting distinct underlying mechanisms. Here, we further tested this dissociation by asking whether an empirical trial-by-trial learning rate (calculated using the same approach as Chapter 1; see Figure 1.2) modulated memory for the outcome event that produced the RPE, or whether learning rate influenced memory for the subsequent cue event where the value update was applied. We did not find learning rates to predict memory for the outcome events that generated them (β = 0.05, t = 0.98, p = 0.33), nor did we find them to predict memory for the subsequent cue events where the value update occurred (β = -0.05, t = -1.31, p = 0.20). This was even true for change-point events that led to the highest learning rates (analyzed separately, outcome memory as a function of the learning rate generated by that trial: (β = -0.04, t = -0.53, p = 0.60); cue memory as a function of the learning rate applied to that trial: (β = -0.08, t = -1.35, p = 0.18).

3.3 EXPERIMENT 2

Method

Participants

We recruited 200 participants for each condition on MTurk. We followed the same procedure and exclusion criteria as Experiment 1, leading to a sample of 163 participants in Condition 1, and 168 participants in Condition 2 (331 participants in total).

Task design

The experiment followed the same procedure as Experiment 1, with differences listed below. Instead of learning the value of a single category (objects) within two reward contexts (Exp. 1), participants learned the value of two categories (indoor and outdoor scenes) within one reward context, thereby eliciting RPEs at cue as well as at outcome. The average value of one of the scene categories was higher than the other, and average values, as well the variance of values (same for both scene categories; σ = 15.81), remained constant throughout the experiment. In order to test a range of RPEs experienced at cue, participants learned in a reward environment where either (1) the average means of the two scene categories were close to each other (μ-high-variance = 60¢, μ-low-variance = 40¢; 'Condition 1'), or (2) further apart (μ-high-variance = 80¢, μ-low-variance = 20¢; 'Condition 2').

Reward Learning: As before, participants saw two different trial-unique events at reward cue and outcome with the cue and the outcome scenes belonging to the same scene category. Participants completed 30 trials during learning (15 trials for each scene category; 60 trial-unique scenes).

Recognition memory: The memory test included 64 trials (32 old and 32 new images).

Choice: As before, participants chose between scenes that belonged to the same reward pair (8 trials). They also chose between two cue scenes or two outcome scenes that were associated with different reward outcomes (6 trials). This was done to check whether the rewards experienced on each trial led to a consistent preference for the more rewarding scene for both cue and outcome events. Choice results are reported in Section 3.4.

Reinforcement learning results

In this experiment, we compared all described models as participants experienced RPEs at both cue and outcome. We found that the model including all three tested modulators of learning rate—an unsigned RPE at outcome (Pearce-Hall component), a signed RPE at cue (Mackintosh component), and an exponential decay—fit participant value estimates best (Figure 3.4C, Table 3.1). This model had a significantly better (e.g., lower) BIC score than the next best fitting model (RW-PH-D; $t(330) = -5.94$, $p < 0.001$). The likelihood-ratio test comparing these two models further confirmed a significantly better fit for the full model ($\chi^2(331) = 808.60$, $p < 0.001$).

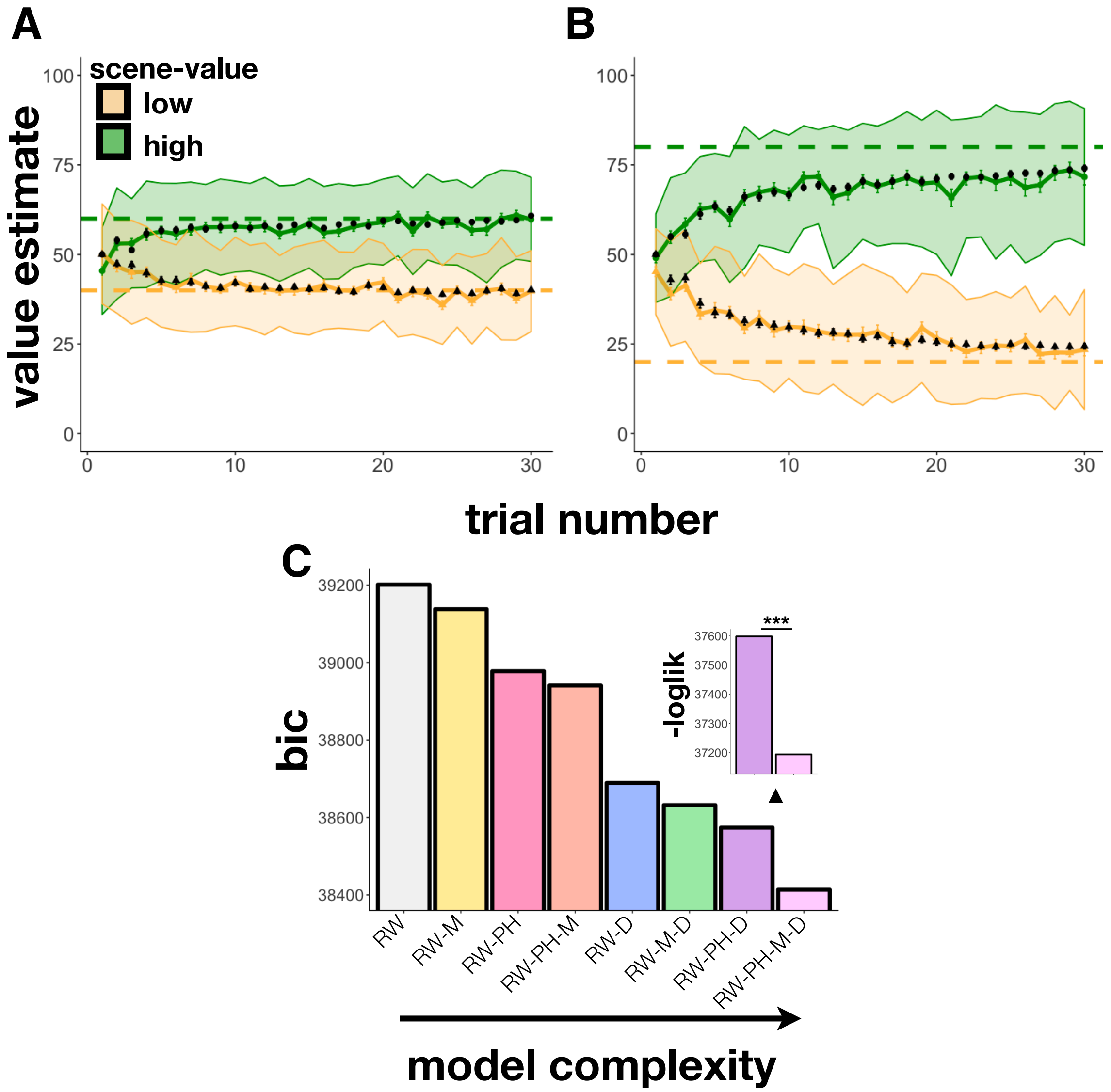


Figure 3.4: Experiment 2 learning behavior and modeling results. **A-B**. Experiment 2 average participant value estimates as a function of trial number in Condition 1 (A) and Condition 2 (B), and average model value predictions of the RW-PH-M-D model in black. True rewards not shown as they varied across subjects. **C**. Total BIC scores across subjects for models tested in Experiment 2. Lower scores indicate better fit; insets show negative log-likelihoods for the two best fitting models, as tested in the likelihood-ratio test. The RW-PH-M-D model significantly outperformed the next best fitting model according to both BIC and likelihood-ratio test.

model	parameters	-LL	BIC
RW	α	20911.33	20983.26
		38957.07	39201.04
RW-PH	η, κ	20733.54 (*d* = -177.79)	20877.41 (*d* = -105.86)
		38489.88 (*d* = -467.20)	38977.82 (*d* = -223.23)
RW-M	η, γ	38650.04 (*d* = -307.04)	39137.98 (*d* = -63.07)
RW-D	η, N, λ	20754.96 (*d* = -156.37)	20970.76 (*d* = -12.50)
		37957.44 (*d* = -999.63)	38689.35 (*d* = -511.69)
RW-PH-M	η, κ, γ	38208.66 (*d* = -748.41)	38940.57 (*d* = -260.47)
RW-PH-D	η, κ, N, λ	20627.11 (*d* = -284.22)	20914.84 (*d* = -68.42)
		37598.08 (*d* = -1358.99)	38573.96 (*d* = -627.08)
RW-M-D	η, γ, N, λ	37655.84 (*d* = -1301.23)	38631.72 (*d* = -569.32)
RW-PH-M-D	η, κ, γ, N, λ	37193.78 (*d* = -1763.29)	38413.63 (*d* = -787.41)

Table 3.1: Model parameters and fit results. 'RW': Rescorla-Wagner, 'PH': Pearce-Hall, 'M': Mackintosh, 'D': Decay. Negative log-likelihood and BIC scores across participants for Experiment 1 (gray background) and Experiment 2; "*d*" refers to the difference in score between the tested model and the baseline model ('RW'). Lower scores indicate better fit. In Experiment 1, models that include a PH modulation of learning rate performed better, and in Experiment 2, the model that includes all three tested components of learning rate (PH, M, D) performed the best.

Memory results

Memory by learning condition

Here, we hypothesized that increased learning for two reward categories (eliciting RPEs at cue), should lead to attentional enhancement and improved encoding of cue events. This is in contrast to Experiment 1 where, in the absence of cue RPEs, we did not observe learning to modulate cue memory. First, we again found better memory for cue events in both conditions of the experiment (Condition

1: β = -0.53, t = -7.34, p < 0.001; Condition 2: β = -0.66, t = -8.11, p < 0.001; Figure 3.5A-B). However, unlike Experiment 1, memory for cue events increased as a function of trial number, relative to memory for outcome events, in both conditions (Condition 1: cue-outcome memory interaction β = -0.25, t = -3.89, p < 0.001; Condition 2: cue-outcome memory interaction β = -0.20, t = -3.29, p < 0.001). In fact, at the beginning of learning, memory for cue events was not better than memory for outcome events, and the difference emerged as a function of learning. The two conditions varied in learning difficulty, due to the larger overlap between the values of the two scene categories in Condition 1.

Accordingly, in Condition 2, better memory for cue events was already evident at the beginning of learning, leading to a less steep increase in memory for cue events than in Condition 1, as assessed by a difference in the slope of this effect (μ-$cond_1$-slope = 0.25, μ-$cond_2$-slope= 0.17, $t(328.89)$ = 5.68, p < 0.001). These results suggest that larger differences between reward expectations for the two scene categories enhanced memory for cue events, as evidenced by an increase in cue memory over learning, which furthermore occurred earlier in an easier learning environment. See below, 'Memory by RPE signals', for a direct test of whether reward expectations, as represented by a cue RPE (i.e., a relative value signal), influenced memory for cue events.

Memory by RPE signals

In this experiment, we tested both outcome (as in Exp. 1) and cue RPEs on memory. When testing the effect of signed and unsigned RPE signals on memory for scenes, we found that signed cue RPEs enhanced memory for both cue and outcome scenes (mixed-effects linear regression β = 0.08, t = 3.43, p < 0.001; model controlled for differences in memory for cue and outcome events), meaning that memory for more valued scenes was boosted relative to less valued scenes. Furthermore, we found a separate effect of unsigned cue RPE on memory, such that the more participants had separated the values of the two scene categories (i.e., the more they had learned), the better their memory for either scene category (β = 0.07, t = 2.49, p = 0.01). Here, we did not find an overall effect of unsigned or signed outcome RPEs on memory (|outcome RPE|: β = 0.04, t = 1.50, p = 0.14; outcome RPE: β = -0.04, t = -1.61, p = 0.11). However, when testing for the interaction of unsigned outcome RPEs on cue versus outcome memory in this model, we did find a significant effect ($\beta = 0.09$, $t = 2.06$, $p = 0.04$), showing that unsigned outcome RPEs modulated memory, as in Experiment 1, but only for outcome events. We further characterize this effect below.

When we separately modeled the effects of cue and outcome RPE signals on cue and outcome memory, we found that signed cue RPEs boosted memory for both cue and outcome events, and especially for cue events (cue memory as a function of cue RPE: M = 0.08, HDI [0.01, 0.15], $\beta = 0.08$, $t = 2.64$, $p = 0.008$, outcome memory as a function of cue RPE: M = 0.07, HDI [-0.01, 0.12], $\beta = 0.07$, $t = 2.26$, $p = 0.03$; Figure 3.5C). The unsigned cue RPE further modulated cue memory, while this effect was trending for outcome memory (cue memory as a function of |cue RPE|: M = 0.09, HDI [0.01, 0.16], $\beta = 0.07$, $t = 2.00$, $p = 0.05$, outcome memory as a function of |cue RPE|: M = 0.06, HDI [-0.01, 0.13], $\beta = 0.07$, $t = 1.82$, $p = 0.07$; Figure 3.5D).

On the other hand, unsigned outcome RPEs only enhanced memory for outcome events (outcome memory as a function of |outcome RPE|: M = 0.14, HDI [0.05, 0.23], $\beta = 0.09$, $t = 2.59$, $p = 0.009$; cue memory as a function of |outcome RPE|: M = -0.02, HDI [-0.11, 0.06], $\beta = -0.01$, $t = -0.46$, $p = 0.64$; Figure 3.5E). We again did not find any effect of signed outcome RPEs on memory for either cue or outcome events (outcome memory as a function of outcome RPE: M = -0.03, HDI [-0.10, 0.03], $\beta = -0.05$, $t = -1.46$, $p = 0.15$, Figure 3.5F; cue memory as a function of outcome RPE: M = 0.002, HDI [-0.07, 0.07], $\beta = -0.02$, $t = -0.66$, $p = 0.51$).

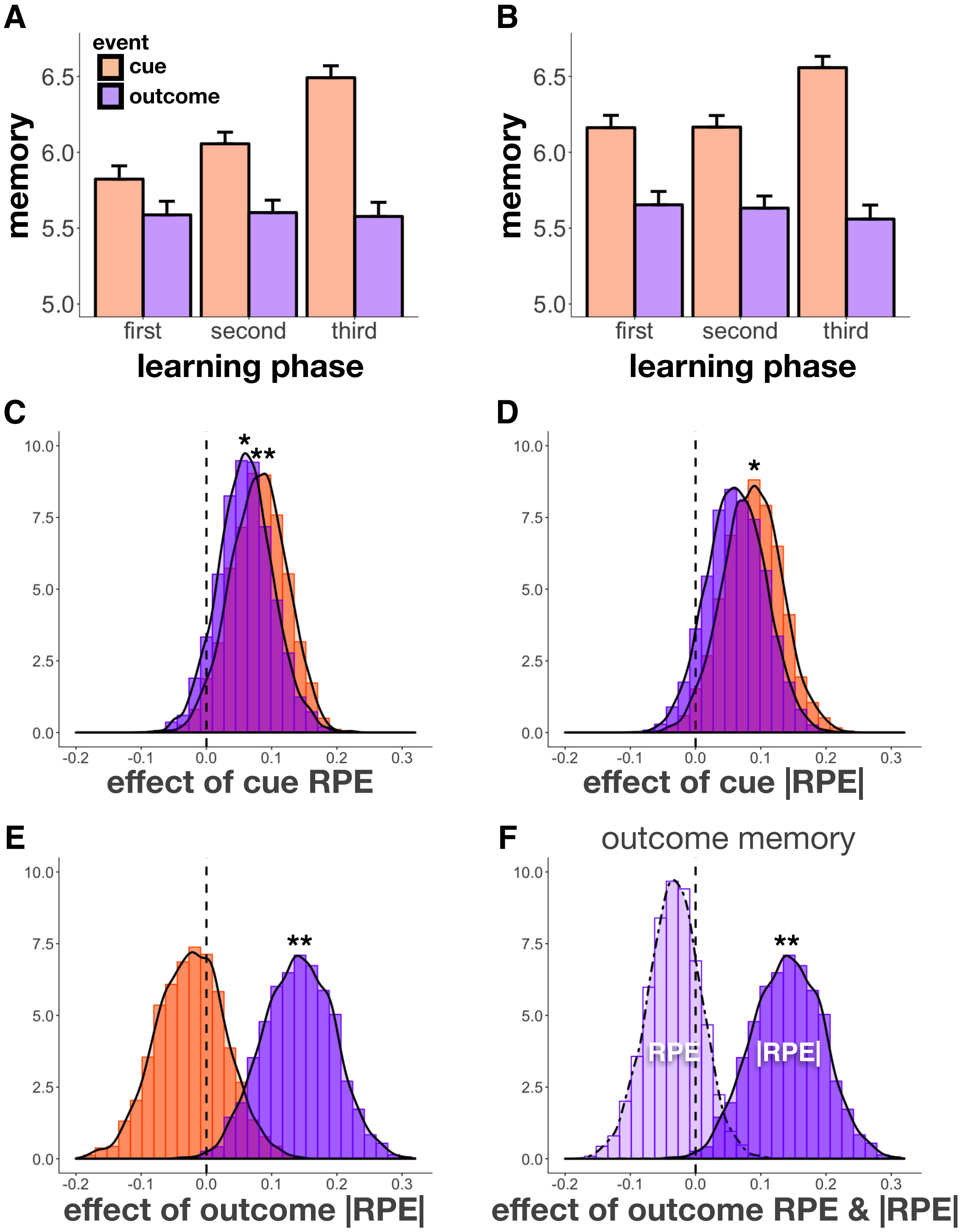


Figure 3.5: Experiment 2 memory results. **A**. Experiment 2 memory score in Condition 1; **B**. Experiment 2 memory score in Condition 2, both as a function of 'learning phase' (learning trial number binned in thirds). In both conditions, learning enhances cue memory. Furthermore, the slope of improvement in cue memory is less steep in Condition 2, where an easier learning environment leads to earlier enhancement of cue versus outcome memory. **C-F**. Parameter distributions from hierarchical models of memory. (C) Signed cue RPE effects on cue and outcome memory. Signed cue RPEs boosted memory for both the item presented at cue and the item presented at outcome, with a stronger effect for cue memory; (D) Unsigned cue RPE effects on cue and outcome memory. Unsigned cue RPEs additionally, and separately, enhanced memory for cue events but not for outcome events; (E) Unsigned outcome RPE effect on cue and outcome memory. Different from the unsigned cue RPEs, unsigned outcome RPEs only increased memory for outcome events; (F) RPE effects on outcome memory. Signed outcome RPEs again did not influence memory (unsigned RPE plot identical to panel E, repeated for comparison with signed RPE).

Memory by learning rate

As in Experiment 1, we did not find learning rates to influence memory for the outcome events that generated them ($\beta = 0.01$, $t = 0.34$, $p = 0.73$), nor did we find learning rates to affect memory for the cue events where the value update was applied ($\beta = 0.05$, $t = 1.42$, $p = 0.16$). Echoing our findings in Chapter 1, we did not find an association between learning rate for values and memory for those events, suggesting separate mechanisms.

3.4 CHOICE RESULTS

In both experiments, at the end of the experiment, we asked participants to choose between previously-encountered items that had (1) belonged to different learning trials and were either both cue or outcome events, or (2) belonged to the same trial (i.e., the cue and the outcome event for that trial), and were thus associated with the same value estimate and the same reward. In this phase, no new outcomes were presented.

(1) Choices across trials: In Experiment 1, we investigated whether participants were biased to prefer cues or outcomes that were from different variance contexts (i.e., were associated with different levels of outcome RPEs), but linked to similar reward outcomes. We did not find variance context, nor unsigned outcome RPE, to modulate preference (variance: $\beta = 0.29$, $z = 1.57$, $p = 0.10$; $|$outcome RPE$|$: $\beta = 0.05$, $z = 0.49$, $p = 0.62$). In Experiment 2, participants chose between cue and outcome pairs that were associated with different reward outcomes. Participants chose both cues and outcomes that were associated with higher rewards, confirming that they had associated the reward outcome with both cue and outcome events (condition 1: $\beta = 1.00$, $z = 8.50$, $p < 0.001$; condition 2: $\beta = 2.79$, $z = 10.96$, $p < 0.001$; Figure 3.6A). Controlling for this effect of reward outcome, we also found that participants chose both cue and outcome events associated with larger cue RPEs, i.e., the more participants had valued the scene category at the time of encoding, the more likely they were to choose those cue and outcome events at the choice test (condition 1, cue RPE: $\beta = 1.04$, $z = 5.94$, $p < 0.001$, reward: $\beta = 0.39$, $z = 3.28$, $p = 0.001$; condition 2, cue RPE: $\beta = 0.83$, $z = 3.88$, $p < 0.001$, reward: $\beta = 1.38$, $z = 6.53$, $p < 0.001$; Figure 3.6B).

(2) Choices within trial: The above results show that participants associated both cue and outcome events with reward outcome as well as their value for that scene category during encoding. Nevertheless, we also asked participants to choose between cue and outcome events from the same trial, which had the same associated value and reward outcome. The only putative learning component that differentiated these two events was therefore the RPE that participants experienced at outcome. Here, we found (Exp. 1), and replicated (Exp. 2), an effect whereby the greater the (signed) RPE experienced on that trial at outcome, the more likely participants were to prefer the outcome event over the cue event (Exp. 1: $\beta = 0.27$, $z = 3.27$, $p = 0.001$, Figure 3.6C; Exp. 2: condition 1: $\beta = 0.28$, $z = 4.18$, $p < 0.001$, condition 2: $\beta = 0.21$, $z = 3.67$, $p < 0.001$, Figure 3.6D).

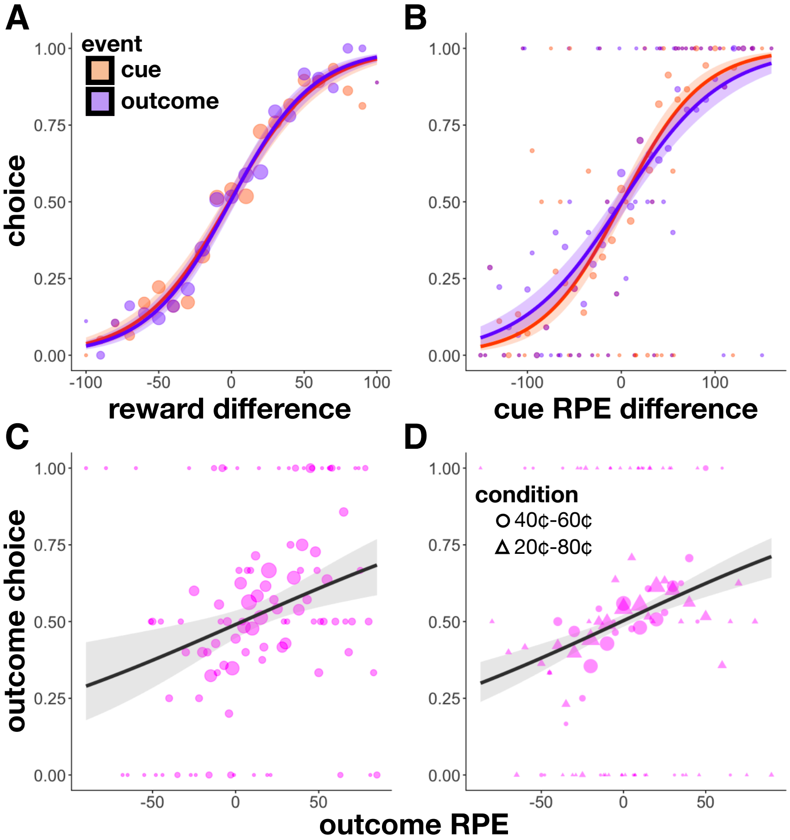


Figure 3.6: Choice results. **A**. Experiment 2: Choice probability as a function of the difference in reward outcomes between two cue or two outcome events. Participants were more likely to choose cue and outcome events that had been associated with higher reward outcomes. **B**. Experiment 2: Choice probability as a function of the difference in cue RPE between two cue or two outcome events. Participants were more likely to choose cue and outcome events that they had associated with a more valuable scene category (relative to the other scene category) at the time of encoding. **C**-**D**. Choice between a cue and an outcome event from the same trial (same associated reward outcome and cue RPE) in Experiment 1 (C) and 2 (D). Participants were more likely to prefer the outcome event if it had been associated with a higher (signed) outcome RPE. Size of the points reflects the size of that sample. Choice was fit using a logistic function, and shaded regions reflect 95% confidence intervals.

3.5 Discussion

In two experiments, we found that distinct reward prediction error (RPE) signals, one occurring at cue and one at outcome, dynamically influenced learning and memory for those events. Drawing on classic associative models of attention (Pearce & Mackintosh, 2010), we found that an unsigned RPE at reward outcome, consistent with a Pearce-Hall model of learning (Pearce & Hall, 1980), and a signed RPE at reward cue, consistent with a Mackintosh model of learning (Mackintosh, 1975), modulated a dynamic learning rate in reinforcement learning models that predicted behavior better than models without those attentional components.

RPE signals at cue and outcome also enhanced memory for associated events. In Experiment 1, participants learned the value of a single reward category while experiencing large unsigned RPEs at outcome due to high (versus low) levels of outcome variance and unexpected changes in the mean of the underlying reward distribution ('change-points'). Here, unsigned RPEs at reward outcome increased memory for scenes accompanying both the cue and outcome, and in particular the latter. In Experiment 2, participants learned the value of two reward categories (designated by indoor and outdoor scenes), which meant that they experienced RPEs both at the time of the cue (as they could not predict which category would be offered on any given trial) and at the time of the reward outcome. Unlike Experiment 1 where memory for the cue event remained relatively stable throughout learning, in Experiment 2, memory for the cue event (but not the outcome event) increased throughout learning. This increase was supported by the gradual buildup of a signed RPE at cue, which enhanced memory for more valued reward cues (and to a lesser extent, their outcome events), as well as an unsigned RPE at cue that benefited memory for both reward cues the more participants had separated the values of the two reward categories (i.e., the more the participants had learned). We again found unsigned RPEs at reward outcome to boost memory for outcome events, but here we did not find them to influence memory for cue events.

We therefore found signed and unsigned RPEs to modulate a dynamic learning rate and enhance memory for those events, effects which we argue here to be both, at least, initiated by the same attentional mechanisms associated with those RPEs. Nevertheless, we did not find learning rate to predict memory for either the outcome event where the update was signalled nor for the cue event where the update was applied. Consistent with our Chapter 1 results, we maintain that although RPEs influence both learning rate for values and memory of those events, they, in part, rely on separate mechanisms. For a fuller discussion of this point, see 'Distinct unsigned-RPE effects on learning and memory' in the Conclusion (Chapter 5).

Last, in a final choice test, participants preferred both cue and outcome scenes that had been associated with higher reward outcomes and more valued scene categories; however, when choosing between the cue and outcome scenes of a single trial (i.e., two scenes with the same associated reward and value), higher signed RPEs at outcome, which we did not find to modulate memory, led to greater 'irrational' preference for the outcome event.

Reward prediction errors modulate a dynamic learning rate

We compared different reinforcement learning models that included the contribution of attentional components in modulating a dynamic learning rate. Note that our experiments depart from classic paradigms that investigate the role of selective attention in learning, because we did not investigate the relative allocation of attention (or learning resources) between competing stimuli presented simultaneously. Instead, we presented one stimulus at a time. Nevertheless, our data allowed us to model and test the amount of learning for each stimulus, and investigate its relationship to RPEs. We tested the influence of a Pearce-Hall attention component (Pearce & Hall, 1980), where the unsigned RPE at reward outcome enhances attention, as modeled by an increase in learning rate. We found that allowing this signal to update the trial-by-trial learning rate led to better predictions of learning behavior.

We also tested the influence of a Mackintosh attention component (Mackintosh, 1975), which contrary to the Pearce-Hall model, predicts an increase in attention (and learning rate) for more valuable and predictive cues. We modeled the Mackintosh signal as the (signed) difference in learned value between the reward-predicting cues, which we referred to as a signed RPE at cue. In other words, the more one cue was valued over the other, the stronger (and more positive) the associative strength. It may be worth mentioning that an unsigned RPE at cue could similarly reflect a Mackintosh signal as larger distances between values indicate greater learned predictiveness. Nevertheless, we found that modulating learning rate according to a signed-RPE signal at cue also predicted behavior better than a model that did not include this component. This was in addition to the separate contribution of the Pearce-Hall signal to learning rate.

Signed and unsigned reward prediction errors enhance memory

RPEs modulated not only learning rate, but also memory for trial-unique items associated with cues and outcomes. In particular, we replicated previous results showing better memory for items that were accompanied by high-magnitude RPEs at reward outcome, either due to outcome variance (Chapter 1; Rouhani et al., 2018) or reward change-points (Chapter 4; Rouhani et al., 2020). Furthermore, we found this mnemonic benefit to be particularly strong for events experienced at outcome. Although unsigned outcome RPEs did enhance memory for cue events in Experiment 1, this effect was weaker than that for outcome events. Moreover, in Experiment 2, where RPEs were additionally experienced at cue, we no longer saw any influence of the unsigned outcome RPE signal on cue memory. We therefore hypothesize that increased attention brought on by large unsigned RPEs during reward outcome engages the locus coeruleus (LC), which co-releases norepinephrine and dopamine to modulate hippocampal plasticity (Kempadoo et al., 2016; Takeuchi et al., 2016; Wagatsuma et al., 2017).

As in prior work (Rouhani et al., 2018, 2020), we did not find signed RPEs at reward outcome to modulate memory for any event, but we note that such an effect has been demonstrated in paradigms outside of reinforcement learning (Marvin & Shohamy, 2016; Ergo, De Loof, Janssens, & Verguts, 2019; De Loof et al., 2018) and for adolescents, but not adults, in reinforcement learning (Davidow et al., 2016). Instead, we found that the signed RPE at cue enhanced memory for both cue and outcome events. That is, as learning progressed, cues that were more valuable (and therefore elicited a larger signed RPE at cue) were associated with better memory for both cue and outcome scenes. This effect was stronger in enhancing cue memory, consistent with previous work showing better memory for cues associated with higher rewards, prior to receiving the actual outcome (Stanek et al., 2019; Jang et al., 2019). We speculate this signed RPE effect on memory to be supported by the (signed) dopaminergic RPE moving from reward outcome to the cue predicting reward during learning (Barto, 1995; Montague et al., 1996; Schultz et al., 1997).

In Experiment 2, we in fact saw an incremental increase in memory for cue events throughout learning, but not for outcome events. This increase, along with the signed cue RPE described above, was supported by an unsigned RPE at cue, improving memory for both high-value cues and low-value cues as participants learned to separate the values of the two reward categories. Although we consider the signed RPE at cue to reflect a Mackintosh-type (Mackintosh, 1975) attention signal, in our paradigm, larger unsigned RPEs at cue also demonstrated stronger learned associative strengths for each reward category, further bolstering an account whereby greater reward predictiveness leads to putative attentional enhancement that strengthens encoding of those cue events. Furthermore, both conditions of Experiment 2 showed this improvement in cue memory with more learning; however, in the easier learning condition, better memory for cue events appeared earlier, and the slope of memory improvement over time was less steep than the more difficult learning condition, again reflecting that learning is modulating cue memory.

In Experiment 1, however, where there were no RPEs experienced at cue, cue memory was higher than outcome memory throughout learning. This consistent better memory for cue versus outcome items is perhaps unsurprising given that during outcome events, the monetary outcome itself may distract from encoding the outcome event. Nevertheless, these results suggest that learning about more than one reward category (thereby creating cue RPEs) may be important in generating learning-driven changes to cue memory.

INTERACTIONS BETWEEN REINFORCEMENT LEARNING AND MEMORY SYSTEMS

Although we did not measure neural activity in this study, distinguishing the mnemonic effects of signed and unsigned RPEs in the brain may be fruitful in characterizing two distinct memory mechanisms. As noted above, signed RPEs increase dopaminergic firing from the midbrain (Barto, 1995; Montague et al., 1996; Schultz et al., 1997), whereas unsigned RPEs increase noradrenergic (as well as dopaminergic) firing from the LC (Takeuchi et al., 2016; Kempadoo et al., 2016; Wagatsuma et al., 2017). Recent work makes predictions about how these distinct mechanisms may differentially influence memory (Duszkiewicz et al., 2018; Clewett & Murty, 2019; Hauser et al., 2019). Midbrain dopamine is thought to initiate 'behavioral activation' (Clewett & Murty, 2019), such as increased vigor during periods of reward anticipation, which promotes the integration of higher-order representations, like value formation, giving rise to semantic memories (Duszkiewicz et al., 2018). The LC-norepinephrine system, on the other hand, promotes selectivity for salient events such as surprising outcomes, giving rise to distinctive, episodic memories (Duszkiewicz et al., 2018). How these RPE signals may act on different kinds of memory is an important avenue for future research.

In this paradigm, we did not dissociate the effects of cue RPE versus reward anticipation on memory (for an experiment that does this, see Stanek et al., 2019). However, we predict that phasic signed RPEs at cue would initiate and enhance a more sustained, and potentially ramping, period of reward anticipation, leading to memory benefits regardless of the exact timepoint following the cue. In fact, Iigaya et al. recently offered a computational model whereby RPEs amplify anticipatory value (i.e., the 'pleasure of savoring'; Iigaya et al., 2019). They further characterize a neural circuit whereby the hippocampus – tracking unsigned RPEs at outcome – enhances the functional coupling between the dopaminergic midbrain (encoding the signed RPE at outcome) and the ventromedial prefrontal cortex (encoding anticipatory value) to boost reward anticipation. The authors speculate that the cognitive imagining of future rewards may drive hippocampal orchestration of reward anticipation. It is, however, unclear whether hippocampal activation here reflects greater engagement in retrieval processes (supporting the simulation of future rewards) or in encoding processes, consistent with previous work showing better memory for events experienced during reward anticipation (Stanek et al., 2019; Murty & Adcock, 2014; Wittmann et al., 2005). Future work should identify the dynamics of hippocampal encoding and retrieval states (Hasselmo, Bodelón, & Wyble, 2002; Duncan et al., 2012) over the period of reward anticipation.

In our experiments, we found a collaborative interaction between reinforcement learning and episodic memory systems: more rewarding cues and more surprising outcomes were prioritized in memory, thereby promoting adaptive behavior. Nonetheless, in other paradigms, these two systems have been shown to compete for processing resources: compromised feedback-based learning has been associated with enhanced episodic memory, both behaviorally and neurally (Foerde et al., 2012; Wimmer et al., 2014). In fact, Wimmer et al. showed that better memory for reward-predicting cues was associated with weaker striatal RPEs at reward outcome. In our experiments, we did not find any effects of signed RPEs at reward outcome on memory (Wimmer et al., 2014). However, there are several notable differences between this prior task and ours: we tested Pavlovian learning, not instrumental learning; we presented one cue at a time, rather than two cues on every trial; and we tested for memory immediately after learning, rather than 24 hours later (our test did not reflect consolidation effects). Future studies should investigate which learning conditions (e.g. Pavlovian versus instrumental) engage more collaborative versus competitive interactions between reinforcement learning and episodic memory systems.

Positive reward prediction errors bias preference

At the end of our experiments, we investigated how RPE signals influence subsequent choice. Participants chose both cue and outcome events that had been associated with higher reward outcomes as well as higher (relative) value of a scene category. This confirmed that participants were following instructions: they were associating both the cue and outcome scenes with the value of that scene category as well as with the specific reward outcome on that trial. Interestingly, when asked to choose between a cue and an outcome item from the same trial (i.e., where there should be no preference for either item), we found and replicated an effect (in both Experiments 1 and 2) whereby the higher the signed RPE at outcome, the more participants preferred the outcome event. Therefore, although signed RPEs at outcome did not modulate memory, they did predict subsequent choice, pointing to a hedonic component of the signed RPE in shaping preferences. This finding is consistent with work maintaining that RPEs drive changes in emotional or affective states (Villano, Otto, Chiemeka, Gillis, & Heller, 2020; Eldar & Niv, 2015; Eldar, Rutledge, Dolan, & Niv, 2016; Rutledge, Skandali, Dayan, & Dolan, 2014), and we propose that this putative change in affect biased preference for the associated outcome event.

To conclude, in two experiments, we showed that reward prediction errors (RPEs) generated by a reward-predicting cue and a reward outcome modulated learning rate during reinforcement learning, in line with classic attentional models of learning. These signals further enhanced memory for events associated with larger unsigned RPEs experienced at outcome and larger signed RPEs experienced at cue. Moreover, although a signed RPE at outcome did not predict memory, it did 'irrationally' bias preference for that outcome event. Our findings highlight the interaction of prediction errors, potentially signaled by midbrain dopamine and locus-coeruleus norepinephrine, with mnemonic processes.

4

Reward Prediction Errors Create Event Boundaries in Memory[*]

In previous chapters, we showed that larger positive or negative RPEs experienced during reward learning lead to improved memory for those surprising events (Rouhani et al., 2018). However, the mechanism behind this enhanced memory is unclear. For example, imagine you watch a new episode of what had long been your favorite television show, only to find that you strongly dislike it. Worse, this bad episode indicates a decrease in the show's quality (e.g. brought on by a change in writers). Is the episode where the quality of the show changed better remembered because it is more strongly stamped in memory? Or is it better remembered because it predicts a meaningful change in the state of the show, thereby separating the pleasant episodes that came before it from the unpleasant episodes that followed, creating separate clusters in memory? In other words, do high RPEs lead to better memory because they bind events more strongly to the context in which the event occurred leading to greater accessibility of that memory when cued with context, or because they lead to the creation of a new context, thereby reducing interference from memories that came before?

If high RPEs create a new latent state or context, then we predicted they would act as event boundaries in memory. In fact, prediction errors (outside of the reward domain) are thought to create event boundaries by segmenting the continuous stream of experience into separate memory traces (DuBrow, Rouhani, Niv, & Norman, 2017; Gershman, Radulescu, Norman, & Niv, 2014; Zacks,

---

[*] The content of this chapter is in press in Rouhani, Norman, Niv, & Bornstein, 2020.

Speer, Swallow, Braver, & Reynolds, 2007). It is, however, unknown whether changes in the distribution of rewards, signaled by high RPEs, act as event boundaries in memory. Events boundaries structure the temporal organization of memories by interrupting the integration of events across them. This leads to worse memory for the order of events ('sequence memory') and greater perceived distance for events across rather than within contexts (DuBrow & Davachi, 2013; Horner, Bisby, Wang, Bogus, & Burgess, 2016). This is further predicted by greater representational dissimilarity of those events in the hippocampus (DuBrow & Davachi, 2014; Ezzyat & Davachi, 2014). Interestingly, like high-RPE memories, memory for the event boundary itself is enhanced (Heusser, Ezzyat, Shiff, & Davachi, 2018; Swallow, Zacks, & Abrams, 2009). However, temporal memory for the events across the boundary is worse, suggesting a trade-off between memory for the boundary event and the mnemonic integration of events across the boundary event (Heusser et al., 2018).

In four experiments, we investigated whether latent shifts in the reward distribution of a Pavlovian reinforcement task (which generate high RPEs) create such event boundaries in memory. In all experiments, participants first completed a passive, sequential reward task that included several high RPEs indicating changes in the underlying distribution of rewards. We then investigated the degree to which high RPEs affected the temporal organization of memories through recognition priming as well as sequence and distance memory measures. We reasoned that if high-RPE events are more strongly bound to the context they were encoded in, then events around the high RPE would be more accessible to one another, resulting in improved priming and better sequence memory. On the other hand, if high-RPE events create new contexts in memory, then events that occurred on either side of a high RPE would be less accessible to one another, leading to less effective priming and sequence memory relative to other pairs of events at the same presentation distance.

We further asked, if high RPEs do create event boundaries, where does this boundary occur? In other words, is the high-RPE event the last of the old context or the first of the new one? The latent cause model would predict that, because the RPE event is predictive of the rewards to follow, it should be the first event of a new context (Gershman et al., 2014). However, recent work suggests that event boundaries lead to the neural reinstatement of events that preceded the boundaries (Baldassano et al., 2017; Ben-Yakov & Dudai, 2011; Ben-Yakov, Eshel, & Dudai, 2013; Sols, DuBrow, Davachi, & Fuentemilla, 2017), which could bind the high-RPE event to its predecessors. Here, we characterized where the event boundary occurs by testing for each one of these possibilities. We first tested the associative links between a high-RPE event and its direct predecessor in Experiments 1 and 2, as well as one of the conditions of Experiment 4. However, given the possibility that the high-RPE event is still bound to its predecessor, we next tested whether an event boundary occurs across the high-RPE event, i.e., between the high-RPE event's predecessor and successor, in Experiments 3 and 4.

We used recognition priming (Experiments 1-3) and sequence and distance memory tasks (Experiment 4) to compare associative and temporal memory for high and low-RPE events. We additionally developed a computational model (a variant of the Context Maintenance and Retrieval model (Polyn et al., 2009)), where high RPEs induce mnemonic separation between rewarding events, and used this model to simulate performance on our experiments and test whether it captured our main behavioral results.

OVERVIEW OF EXPERIMENTS

Recognition Priming

In Experiments 1-3, we used a recognition priming task to probe whether RPEs influence the degree to which two sequential events are bound in memory. In recognition priming, recognition for an event is better and faster if it is preceded by the event that occurred before it during encoding (Schwartz, Howard, Jing, & Kahana, 2005; Zwaan, 1996). The idea is that retrieval of an item also reactivates items that were associated with it during encoding, either directly, or indirectly via context, facilitating subsequent recognition of those items. This is strongest for the forward sequence (i.e., each cue will reactivate the subsequent one (Howard & Kahana, 2002). Given this, we reasoned that if a high RPE creates an event boundary that separates the high-RPE event from its predecessor, high-RPE events would become less accessible when primed during retrieval, demonstrating less recognition priming. If, instead, high-RPE events are more strongly bound to the previous event, we would expect the RPE event to be more accessible when primed by the preceding event, leading to enhanced recognition priming. Since evidence of recognition priming is more consistently reported in response latencies rather than memory accuracy (DuBrow & Davachi, 2014; Zwaan, 1996), we used and simulated recognition latency as our measure of recognition priming, but additionally report memory accuracy results.

Sequence and distance memory

In Experiment 4 (and its replication), we further tested whether high-RPE events disrupt the integration of events by probing the temporal ordering and perceived distance between them. Contextual changes (both external and internal to an observer) are thought to increase change in one's internal context, leading to greater perceived time between events (Sahakyan & Smith, 2014). Performance on these measures of temporal memory is modulated by representations in the hippocampus, thought to support the temporal structuring of events in memory (Davachi & Dubrow, 2015): Previous studies have found that greater hippocampal dissimilarity between two events across an event boundary predicts worse sequence memory and larger subjective distances between them (DuBrow & Davachi, 2014, 2016; Ezzyat & Davachi, 2014). For sequence memory, we asked participants to indicate which of two items came first, and for distance memory, we asked participants to indicate how far apart the events had been during encoding. If a high RPE signals an event boundary, we would expect worse sequence memory and greater estimated distances for pairs that include or are interrupted by a high-RPE event. On the other hand, if high-RPE events are more bound to the events around them, thereby activating and compressing the sequence of events in memory, we could expect better sequence memory and shorter estimated distances.

4.1 Experiment 1

Method

Participants

Participants were recruited from Amazon's Mechanical Turk (MTurk), and 35 participants initiated the task (age: 27-67, median = 34; 15 female, 20 male). The sample size chosen was a standard number of pilot subjects to recruit for an MTurk study in our lab. We first obtained informed consent online, and prior to accessing the task, participants had to correctly answer questions that checked for their understanding of the instructions. All procedures were approved by Princeton University's Institutional Review Board. We excluded participants if they (a) missed more than 20 memory trials, or (b) had a memory score of less than 0.5 (memory score was determined by A'; Pollack & Norman, 1964). Using these criteria, we excluded 8 participants, which led to a sample of 27 participants.

Task design

Participants completed 6 blocks, each consisting of learning (36 trials in each block), choice (4 trials in each block), and recognition memory phases (42 trials in each block). In the instructions, participants were told they would be exploring six different 'rooms' (i.e. blocks), defined by distinct color backgrounds, where they would 'find' different photographs and earn 10% of the reward value associated with each photograph. We used a Pavlovian (passive) learning design in order to isolate the effects of changes in reward alone, unconfounded by shifts in responding. In the learning phase, participants passively viewed a sequence of trial-unique images of scenes that were associated with different reward values (Figure 4.1A). On each trial, participants saw the scene image for 1 second, then were shown the image with its associated value for 2 seconds. The individual values of the scenes fluctuated around a fixed mean (means ranged from 10¢ to 90¢ in steps of 10¢). Participants were encouraged to remember the individual values of the photographs as they would be choosing between them later (after each room), and earning the reward value of the chosen image.

In each room, the mean value of the photographs shifted either four or five times. Participants were told that a shift in the mean value of the photographs indicated they had found a new 'collection' of photographs that were more or less valuable than their previous collection. Critically, as a result of these *reward shifts*, participants experienced high positive or negative reward prediction errors whose magnitude ranged from 20¢ to 80¢ (and every 10¢ increment in between; these magnitudes reflect a one-trial difference between current and previous reward). Each participant experienced each magnitude of prediction error 1-2 times, and the number of positive and negative reward shifts was balanced (13 positive and 13 negative high-RPE events across the entire experiment). Within each latent *reward state*, participants experienced at least 5 and at most 9 trials (average = 6.75 trials) where the individual values of the scene images fluctuated around the same mean value (individual reward values never deviated more than 5¢ from the mean value). After learning, within each block, participants completed 4 choice trials that were intended to ensure they paid attention to the values in the passively viewed sequence. On each choice trial, two previously-seen images were presented and the participant chose one, anticipating that the reward value of that image would be added to their payment for participating in the experiment. The 8 images used in the choice test were not used in any other memory test.

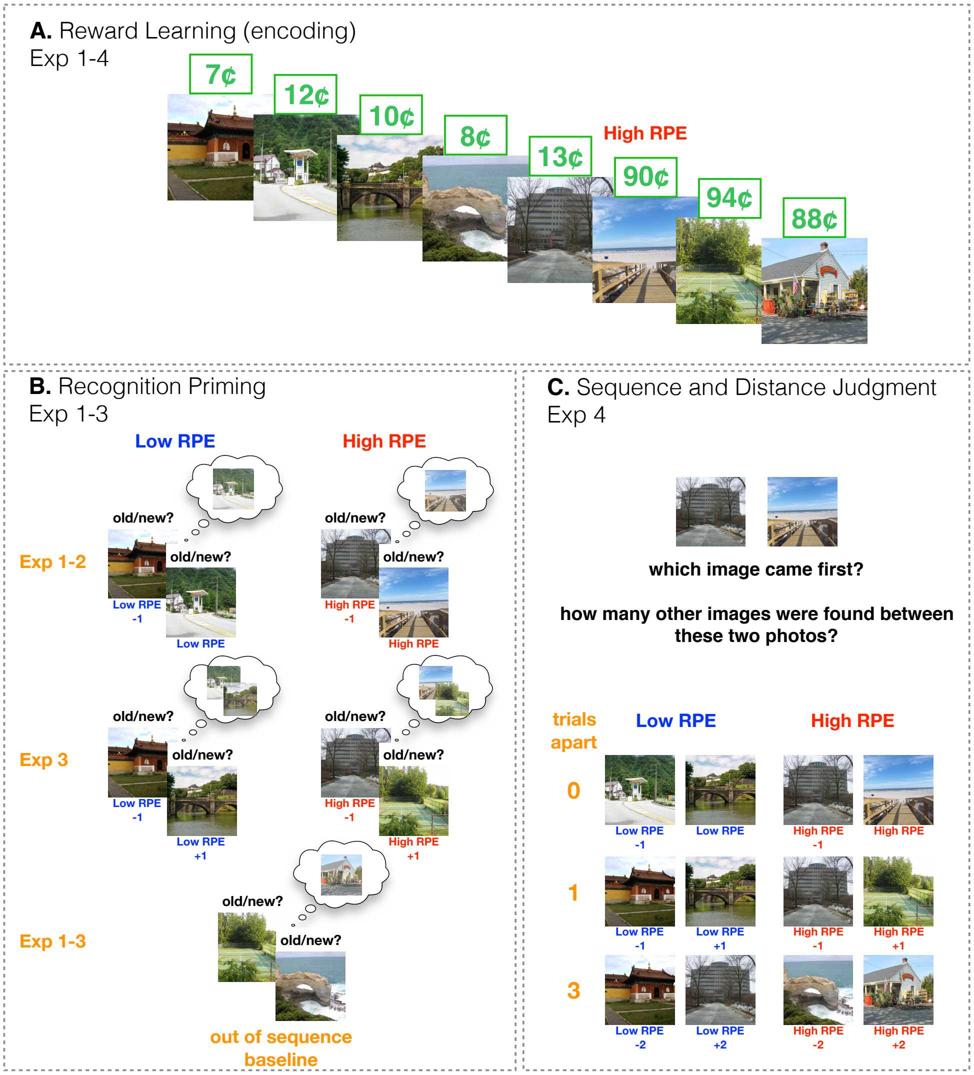


Figure 4.1: Experimental paradigm. A. In all experiments, in each of six blocks, participants first completed a passive reward learning task (the encoding task) where sequences of scenes, each with an associated reward value, were presented. The reward values of the images were contingent on the mean value of the reward state, which shifted 4-5 times each block. B. In Exp. 1-3, after reward learning, participants completed a recognition test where they indicated whether a scene was 'old' or 'new.' We tested for recognition priming of high and low-RPE events, relying on a mechanism by which recognition of an old item (the prime), either directly or indirectly, activates the items that had followed it during encoding (the target), leading to better and faster recognition of target items. Most of the old scenes were presented in pairs that belonged to three different conditions (example stimuli refer to the reward sequence in A): (1) 'low RPE': a pair that was studied consecutively; both items belonged to the same reward state, (2) 'high RPE': a pair that was studied consecutively, however the items belonged to different reward states, (3) 'out of sequence' (baseline): the second item in the test pair actually preceded the first item during encoding (i.e., out of order); the items belonged to different reward states. Recognition priming for low and high-RPE pairs was compared to the out-of sequence pairs. In Exp. 1-2, the low and high-RPE pairs comprised items that were directly one after the other during encoding, whereas in Exp. 3, the pairs were separated by another scene during encoding ('+1'), and so the high-RPE$_{+1}$ pair did not include the high-RPE event itself. C. In Exp. 4, after reward learning, we tested for the temporal memory of two scenes that either belonged to the same reward state (low RPE) or a different reward state (high RPE), and were either 0 (back-to-back), 1 or 3 trials apart. We first asked participants to indicate which of two images came first during encoding (sequence memory), and then for the number of images that occurred between them (distance judgment, scale 0-5). Example pairs (bottom) refer to the reward sequence in A, although unlike the pairs of stimuli presented here, no scene was repeated during testing.

Recognition priming

Following the choice test, we tested for recognition priming of pairs that had either been experienced sequentially during encoding or not (Figure 4.1B). On each recognition trial, participants were asked to indicate 'old' or 'new' for the presented image (by pressing 'o' or 'n' respectively), and to indicate their recognition judgement as quickly as possible. We use 'prime' to refer to the first item and 'target' to refer to the second item in any pair tested during recognition priming. Importantly, the image stayed on screen for 3 seconds regardless of the response time, ensuring that each prime was experienced for the same amount of time.

Recognition trials were comprised of (1) an old scene image ('low- or high-RPE$_{-1}$'), followed by either (a) an old scene image that had followed the prime during learning and belonged to the same reward state ('low RPE'; 4 'priming pairs' within each recognition block, 24 pairs in total), (b) an old scene image that had followed the prime during learning but belonged to a different reward state ('high RPE'; 4 priming pairs within each recognition block, 24 pairs in total); (c) an old scene image that had come before the prime ('out of sequence'; 4 pairs within each recognition block, 24 pairs total); (2) new scene images, representing one-third of the images seen during recognition ('new'; 14 images within each recognition block; 84 images total); (3) 'single' old scene images - half of these items were presented following a new image and the other half after an old image (that had been studied at least 3 trials apart from the tested item) so that participants would not learn to expect old items to always appear in pairs ('single'; 4 images within each recognition block; 24 images total). Note that this pair structure was not disclosed to participants, and all test items were presented as part of one single sequence. The order of the conditions at test was predetermined to minimize unintentional spill-over memory effects during recognition from other old items that appeared close to a test item during learning (although the images themselves were randomized).

Recognition priming can be evidenced by better memory and faster reaction times in recognizing a target item after correctly retrieving the prime, compared to when the target was preceded by an old item that had not preceded it during encoding. Although we report differences in both memory accuracy (hit-rate) and response latencies for target items, recognition priming is more consistently observed in response latencies rather than hit-rates (DuBrow & Davachi, 2014; Zwaan, 1996), and so we focused on characterizing (and modeling) recognition latencies within the above four conditions. We were primarily interested in whether recognition priming was enhanced or interrupted for events that had been associated with a high RPE in comparison to the baseline, which was recognition latency for out-of-sequence targets. The out-of-sequence targets served as the primary baseline for recognition priming since, like the primed high and low-RPE pairs, the first item in the pair is 'old', accounting for any recognition priming effects that would arise from recognizing any old item (Duncan et al., 2012). The single (non-primed) images, where half of the images were preceded by new items, served as an additional baseline (see below, 'Statistical analysis').

If a high-RPE event is bound to the event that occurred immediately before it, we would expect faster recognition of the target than the out-of-sequence target. On the other hand, if high-RPE events create a boundary in memory between the high-RPE event and its predecessor, we would expect similar reaction times in recognizing the high-RPE target and the out-of-sequence target. Together with the RPE condition (high or low), we tested whether the sign of the RPE additionally influenced or interacted with the RPE condition to influence recognition latency. We further examined how well a continuous versus a categorical measure of RPE (high or low) predicted our results.

Statistical analysis

All statistical comparisons were conducted using linear or generalized linear mixed-effects models (using lme4 package in R; Bates et al., 2015), treating participant as a random effect for both the intercept and the slope of the tested fixed effect. To test for differences in memory (i.e., hit-rate) between the primed pairs, we analyzed trials where the prime had been correctly remembered; we did this because of prior research indicating that recognition priming only occurs when the prime is itself remembered (Schwartz et al., 2005). This led to the inclusion of the following number of tested trials in analyzing recognition memory: Experiment 1: mean = 50 (out of 72) trials per participant (range = 36-68 trials), Experiment 2: mean = 47 (out of 66) trials per participant (range = 28-65 trials), Experiment 3: mean = 53 (out of 72) trials per participant (range = 34-71 trials). When testing for recognition priming in reaction time, we analyzed trials where both the prime and the target were correctly remembered. As discussed earlier ('Recognition priming'), this was because we only expected recognition priming to occur for latencies when the prime and the target were correctly remembered. This led to the inclusion of the following number of tested trials in analyzing reaction time: Experiment 1: mean = 37 (out of 72) trials per participant (range = 17-60 trials), Experiment 2: mean = 34 (out of 66) trials per participant (range = 11-60 trials), Experiment 3: mean = 41 (out of 72) trials per participant (range = 14-68 trials). Reaction times were log-transformed and z-scored within participant.

The primary baseline used to assess recognition priming was the out-of-sequence pairs, although the single (non-primed) items can be used as an alternative baseline. The difference between the two conditions was that during the recognition tests, the single items were sometimes preceded by new items and sometimes preceded by old ones (that were not necessarily out of sequence from the item); however, we did not find reliable differences between single items preceded by new versus those preceded by old items within and across all recognition priming experiments. In line with this, the single items were not statistically different than the out-of-sequence targets in recognition latency; we report results for these items as an additional baseline in the following experiments.

Results

Recognition memory

We found that the primed targets were better remembered than the out-of-sequence targets, regardless of the RPE condition (β = 0.35, z = 2.91, p = 0.004, μ-out-of-sequence = 0.68; low RPE: β = 0.38, z = 2.75, p = 0.006, μ = 0.74; high RPE: β = 0.32, z = 2.18, p = 0.03, μ = 0.73; Figure 4.2A). We did not find a difference in memory between the primed high-RPE and low-RPE images (β = -0.04, z = -0.30, p = 0.76) nor between images in the two non-primed conditions (i.e. the out-of-sequence vs. the 'single low RPE' items: β = -0.01, z = -0.05, p = 0.96, μ-single = 0.68).

Recognition latency

The primed targets were more quickly recognized than the out-of-sequence targets (β = -0.15, t = -2.82, p = 0.005, μ-out-of-sequence = -0.05; Figure 4.2B), providing evidence of recognition priming. This was significant for high-RPE targets (β = -0.19, t = -3.11, p = 0.002, μ = -0.24), and trending for low-RPE targets (β = -0.11, t = -1.83, p = 0.06, μ = -0.17). Latencies were moreover no different between the two primed conditions (β = -0.08, t = -1.16, p = 0.25). Additionally, the latencies for correctly recognizing the non-primed targets were not significantly different across conditions (β = 0.02, t = 0.31, p = 0.76, μ-single = -0.04). We did not find an effect of positive versus negative RPE targets (β = 0.04, t = 1.01, p = 0.32), nor did this interact with RPE condition (β = -0.06, t = -0.82, p = 0.41) to influence reaction times. We furthermore did not find that a continuous measure of RPE predicted latencies across RPE conditions (β = -0.0004, t = -0.36, p = 0.72).

Discussion

We found better and faster recognition of items that had been primed, including items that were associated with a high RPE. These results suggested that a high-RPE event is bound to its predecessor. However, given that high-RPE items are generally better remembered (Rouhani et al., 2018), it is possible that the generally stronger memory trace is driving the recognition memory results, and not a stronger association with the previous item. We therefore tested in Experiment 2 whether there are differences in the recognition of primed versus non-primed high-RPE items. Specifically, if a high-RPE event is more bound to the preceding event in memory, then we would expect faster latencies for high-RPE items that are primed versus those that are not primed.

4.2 Experiment 2

Method

Participants

One-hundred participants from MTurk (age: 22-71, median = 35; 46 female, 54 male) were recruited on MTurk. The sample size was chosen because it was the approximate number of subjects needed to detect a medium-sized correlation at 80% power (Hulley, 2007). Following the same exclusion criteria stated in Exp. 1, we excluded 17 participants, leaving a final sample of 83 participants.

Task design

Experiment 2 was identical to Experiment 1 except that during the recognition test we additionally included 'single' (i.e., not primed) scene images associated with high RPEs. We did this to determine whether high-RPE events lead to better and faster recognition because they are more strongly bound to the previous item (and thus show more recognition priming) or because they are more strongly encoded (i.e., a recognition effect not affected by priming). This led to one fewer high-RPE pair within each recognition block, and 41 trials within each recognition block. Across the experiment, for each participant we tested 18 high-RPE priming pairs, 24 low-RPE priming pairs, 24 out-of-sequence pairs, 16 low-RPE single images, and 8 high-RPE single images.

Results

Recognition memory

We again found that the high-RPE primed items were better remembered than the out-of-sequence items ($\beta = 0.24$, $z = 2.34$, $p = 0.02$, μ-high-RPE-primed = 0.76, μ-out-of-sequence = 0.71; Figure 4.2C); however, we did not find them to be better remembered than high-RPE images that were not primed ($\beta = 0.05$, $z = 0.37$, $p = 0.71$, μ-high-RPE-single = 0.78). Therefore, we could not conclude that better recognition memory for the high-RPE images was necessarily a result of recognition priming, further supporting the use of recognition latency, instead of accuracy, as our measure of recognition priming. Additionally, and consistent with previous research (Rouhani et al., 2018), we found high-RPE (single) items were better remembered than low-RPE (single) items ($\beta = 0.39$, $z = 3.40$, $p < 0.001$, μ-low-RPE-single = 0.70).

Recognition latency

We replicated our previous observation of faster reaction times in recognizing the primed high-RPE items than the out-of-sequence ones ($\beta = -0.08$, $t = -2.11$, $p = 0.03$, μ-high-RPE-primed = -0.19, μ-out-of-sequence = -0.10; Figure 4.2D). Importantly, primed high-RPE images were also recognized more rapidly than the non-primed (single) high-RPE images ($\beta = -0.13$, $t = -2.84$, $p = 0.005$, μ-high-RPE-single = -0.06). The faster recognition of the primed high-RPE items thus reflected intact recognition priming. The primed low-RPE targets were not retrieved significantly faster than the out-of-sequence targets ($\beta = -0.05$, $t = 1.32$, $p = 0.19$, μ-low-RPE-primed = -0.15), but were retrieved faster than the non-primed (single) low-RPE images (β = -0.08, t = -2.13, p = 0.03, μ-low-RPE-single = -0.08). When testing for a difference between high and low-RPE targets that were primed versus those that were not primed, we did not find an interaction (β = -0.05, t = -0.76, p = 0.45). Thus, we observed similar levels of recognition priming between high and low-RPE pairs. Again, we did not find an effect of RPE sign (β = 0.02, t = 0.85, p = 0.40), nor did this interact with RPE condition (β = -0.00006, t = -0.001, p = 0.99) to predict reaction times. We similarly did not find that a continuous measure of RPE predicted these latencies across RPE conditions (β = -0.0003, t = -0.48, p = 0.63; for discussion of these results see 'General Discussion').

Discussion

We found that high-RPE items were both better remembered overall, and were also primed (at least with regard to reaction time) by recognition cues. From this, we concluded that high-RPE items were, in fact, linked with the items that had occurred before them during encoding, providing no evidence of an event boundary between a high-RPE event and its predecessor. However, it remained possible that the boundary occurs across rather than during the high-RPE event. To investigate this possibility, we next tested for priming between pairs that had one item in between them during encoding. In other words, we tested for a boundary between the event before and the event after a high RPE. This allowed us to exclude the high-RPE item itself and determine whether we see diminished priming for events across a high-RPE versus those across a low-RPE event.

4.3 EXPERIMENT 3

Method

Participants

We again recruited 100 participants on MTurk (age: 20-66, median = 33.5; 39 female, 61 male), and following the exclusion criteria stated in Experiment 1, we excluded 15 participants, leading to a final sample of 85 participants. The sample size was chosen because it was the approximate number of subjects needed to detect a medium-sized correlation at 80% power (Hulley, 2007).

Task design

The task structure was the same as in Experiments 1 & 2. During recognition, however, instead of testing pairs that had been presented directly one after the other during learning, we tested recognition priming for pairs that had one item in between them during learning. In other words, the high-RPE priming pair never included the high-RPE event itself, allowing us to test whether the events around a high RPE provide evidence of an event boundary. As before, the image immediately preceding the high-RPE event was the prime, but the target was now the image after the high-RPE image ('high RPE_{+1}'). The low-RPE priming pairs had also been one trial apart during learning ('low-RPE_{+1}'), and were selected from the same reward state. All primed targets were therefore associated with low RPEs. We tested 24 high-RPE_{+1} priming pairs, 24 low-RPE_{+1} priming pairs, 24 out-of-sequence pairs, 12 single low-RPE images, and 12 single high-RPE images along with 84 new images, across all 6 blocks of the experiment (42 trials within each recognition block).

Results

Recognition memory

Memory was not significantly better for the primed targets in comparison to the out-of-sequence items ($\beta = 0.13$, $z = 1.57$, $p = 0.12$, μ-out-of-sequence = 0.75), and we did not find a difference between primed high-RPE versus low-RPE events ($\beta = -0.001$, $z = -0.01$, $p = 0.99$, μ-high-RPE-primed = 0.77, μ-low-RPE-primed = 0.77; Figure 4.2E). We again found better memory for high RPE (single) items relative to low-RPE (single) items ($\beta = 0.41$, $z = 3.44$, $p < 0.001$, μ-high-RPE-single = 0.79, μ-low-RPE-single = 0.72).

Recognition latency

When excluding the high-RPE item itself, we no longer observed a recognition priming effect for pairs that spanned a high-RPE event (compared to out-of-sequence, $\beta = -0.02$, $t = -0.53$, $p = 0.60$, μ-high-RPE-primed = -0.15, μ-out-of-sequence = -0.13; Figure 4.2F). We nevertheless did see recognition priming for pairs that spanned a low-RPE event (compared to out-of-sequence, $\beta = -0.08$, $t = -2.42$, $p = 0.02$, μ-low-RPE-primed = -0.21). Moreover, there was now a difference between the latencies of the high- and low- RPE pairs where the high-RPE_{+1} targets were more slowly recognized than the low-RPE_{+1} targets ($\beta = 0.06$, $t = 1.89$, $p = 0.05$). We did not find a signed effect of RPE ($\beta = 0.002$, $t = 0.13$, $p = 0.91$), nor did this interact with RPE condition ($\beta = -0.02$, $t = -0.50$, $p = 0.62$) to influence reaction times. Lastly, we did not find that a continuous measure of RPE predicted latencies across conditions ($\beta = -0.0003$, $t = -0.05$, $p = 0.96$); for discussion of these results, see 'General Discussion').

Discussion

The recognition latency results of Experiment 3 provided evidence that high RPEs serve as an event boundary, and more specifically, that this boundary can be observed for events across a high-RPE event (i.e., between the event before and after a high-RPE) rather than directly between the high-RPE event and its predecessor. The slower latencies in recognizing the item that followed the high-RPE prime, which were now similar to the out-of-sequence pairs and significantly slower than the low-RPE pairs, indicated decreased recognition priming. With this initial evidence of an event boundary, we next tested whether events around a high RPE demonstrate other behavioral markers of event boundaries. For this, we asked whether high-RPE events disrupt the temporal organization of events in memory, leading to worse sequence memory and larger perceived distances between item-pairs that included a high-RPE event versus those that did not.

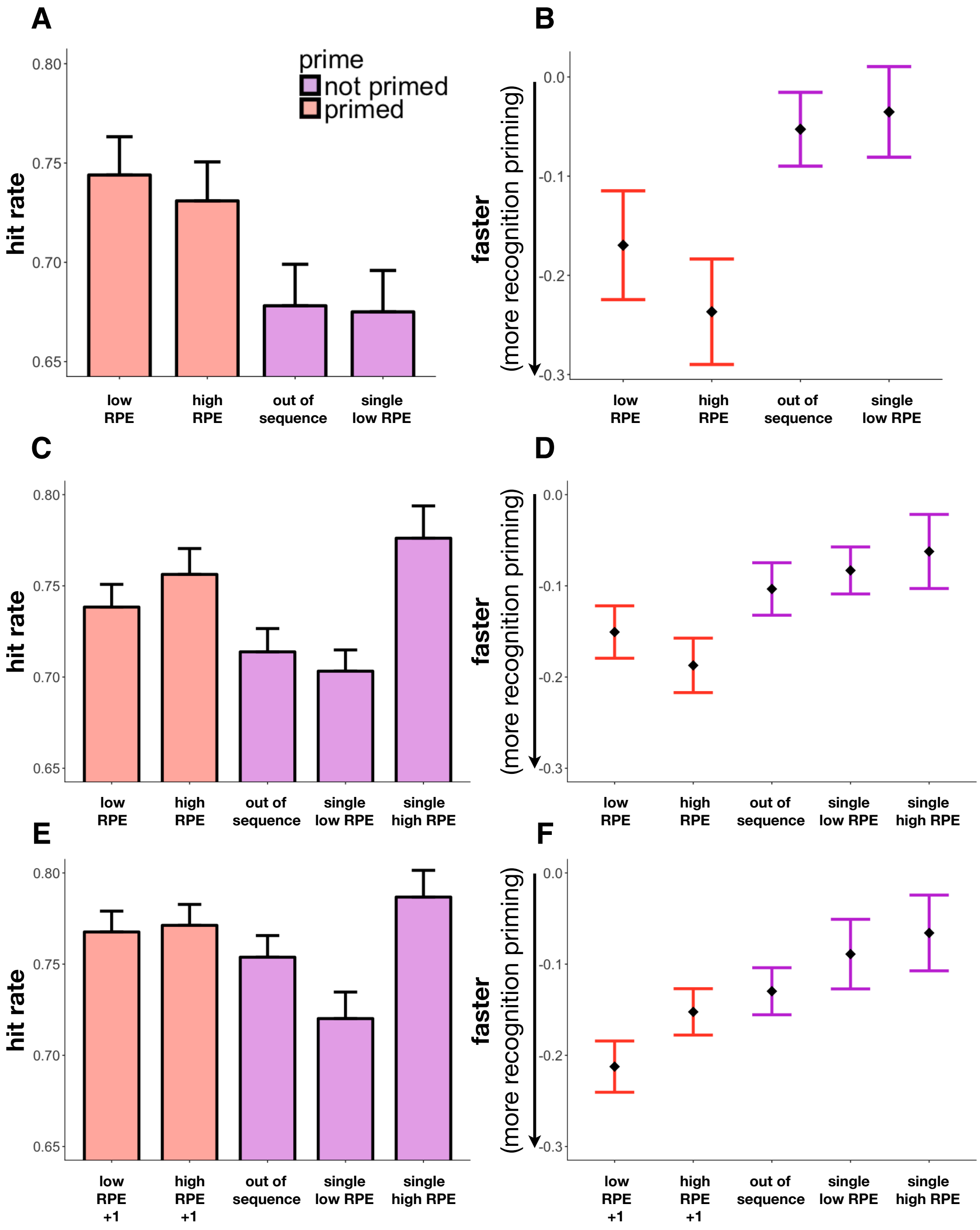


Figure 4.2: Recognition priming results. For paired targets ('low RPE', 'high RPE' and 'out of sequence'), memory is conditioned on correct recognition of the first item in the pair, and response latency is additionally conditioned on correct recognition of the target (i.e., latency is only for 'hits' in all conditions). A. Exp. 1: Recognition memory as a function of item condition. Memory for the sequentially primed targets (low and high RPE) was better than the out-of-sequence and (unpaired) 'single low RPE' targets. B. Exp. 1: Response latencies for correct recognition as a function of item condition. Sequentially primed targets were retrieved faster than items that were not sequentially primed. C. Exp. 2: Recognition memory as a function of item condition. Memory for the primed high-RPE target was no different than the 'single high RPE' target that had not been primed. Thus, memory accuracy did not provide evidence for recognition priming of high-RPE events. D. Exp. 2: Response latencies for correct recognition as a function of item condition. Primed high-RPE targets were retrieved faster than the non-primed high-RPE targets and out-of-sequence targets, thereby demonstrating recognition priming for high-RPE events. E. Exp. 3: Recognition memory as a function of item condition. Primed targets (where the prime was the item presented two trials before the target during encoding) were not remembered better than the out-of-sequence targets. F. Exp. 3: Response latencies for correct recognition as a function of item condition. The high-RPE$_{+1}$ target was no longer retrieved faster than the out-of-sequence target, whereas the low-RPE$_{+1}$ target was still retrieved faster, demonstrating intact recognition priming. Moreover, latencies for the high-RPE$_{+1}$ target were significantly slower than the low-RPE$_{+1}$ target. Error bars represent standard error of the mean (SEM).

4.4 EXPERIMENT 4

Method

Participants

For the first set of this experiment, we recruited 50 participants on MTurk (age: 24-61, median = 38; 26 female, 24 male). We excluded participants if they missed more than 15 trials, which led to the exclusion of 1 participant and a final sample of 49 participants. The sample size chosen was slightly larger than the standard number of pilot subjects given that sequence memory is generally more difficult and noisier than recognition memory.

Subsequently, we ran an additional sample of 80 participants as a pre-registered replication of this experiment (for pre-registration, see Rouhani, 2018). The replication sample size was chosen on the basis of a simulation-based power analysis of the effect seen in the initial sample, which indicated we would have sufficient power (80% probability) of replicating the results with 50 participants. Following common practice of testing around 1.5x the indicated sample size for replication studies, we thus recruited 80 participants on MTurk (age: 24-68, median = 38, 38 female, 42 male), and excluded 3 participants who missed more than 15 trials, leaving a final sample of 77 participants.

Task design

The task structure was the same as in Experiments 1-3; however, instead of testing for recognition memory, here we tested participants' sequence memory and distance judgements for images seen during learning. Worse sequence memory and larger estimated distance between items are considered as evidence of an event boundary in memory (Davachi & Dubrow, 2015). We instructed participants to pay attention to the sequence of images during learning as they would later be asked to order them. After the learning and choice sections in each block, participants were presented with two old scene images on the screen (left/right order counterbalanced), and were asked to indicate which image came first ('sequence memory') and then to estimate how many other images were found between the two (from 0-5; 'distance judgment'; Figure 4.1C). Within each block, participants completed 12 sequence and distance judgment trials. The two scene images either spanned (or even included) a high-RPE event ('high

RPE'; 48 total), or were from the same reward state ('low RPE': 48 total). Additionally, the high/low-RPE manipulation was crossed with a distance manipulation: the pairs had either been presented directly one after the other ('0 between': high-RPE$_{-1}$ and high-RPE events, 24 total), had one item in between them ('1 between': high-RPE$_{-1}$ and high-RPE$_{+1}$ events, 24 total), or had three items in between them ('3 between': high-RPE -2 and high-RPE$_{+2}$ events, 24 total) during learning. Note that the '0 between' high-RPE pairs included the high RPE event and the event that immediately preceded it. The '1 between' high-RPE pairs included the events immediately preceding and following a high-RPE event, and the '3 between' high-RPE pairs included the second event before and the second event after the high-RPE event.

Results

Sequence memory

We found better sequence memory for pairs within the same reward state than across a high RPE ($\beta = 0.25$, $z = 3.46$, $p = 0.0005$; Figure 4.3A), and replicated this main effect in the second sample ($\beta = 0.17$, $z = 2.97$, $p = 0.003$; Figure 4.3B). Interestingly, for the pair that included the high-RPE event itself and its predecessor ('0 between': high RPE$_{-1}$ and high RPE), there was no difference in sequence memory between the pair types (first set: $\beta = -0.03$, $z = -0.26$, $p = 0.79$, μ-high-RPE = 0.52, μ-low-RPE = 0.53; replication set: $\beta = -0.05$, $z = -0.52$, $p = 0.61$, μ-high-RPE = 0.53, μ-low-RPE = 0.54). The difference in sequence memory was instead carried by pairs that were across the high-RPE event, i.e. the pairs that had 1 item in between them (high RPE$_{-1}$ and high RPE$_{+1}$; first set: $\beta = -0.35$, $z = -2.79$, $p = 0.005$, μ-high-RPE = 0.48, μ-low-RPE = 0.56; replication set: $\beta = -0.19$, $z = -1.87$, $p = 0.06$, μ-high-RPE = 0.50, μ-low-RPE = 0.54), and 3 items between them (high RPE$_{-2}$ and high RPE$_{+2}$; first set: $\beta = -0.36$, $z = -2.95$, $p = 0.003$, μ-high-RPE = 0.51, μ-low-RPE = 0.59; replication set: $\beta = -0.27$, $z = -2.88$, $p = 0.004$, μ-high-RPE = 0.52, μ-low-RPE = 0.59).

We also found that for low-RPE pairs, greater distance between items predicted better sequence memory (first set: $\beta = 0.09$, $z = 2.31$, $p = 0.02$; replication set: $\beta = 0.07$, $z = 2.32$, $p = 0.02$), whereas this was not true for high-RPE pairs (first set: $\beta = -0.005$, $z = -0.15$, $p = 0.88$; replication set: $\beta = 0.002$, $z = 0.06$, $p = 0.95$). Although the interaction between distance and high/low RPE was not significant when analyzing each set alone (first set: $\beta = -0.10$, $z = -1.71$, $p = 0.09$; replication set: $\beta = -0.07$, $z = -1.63$, $p = 0.10$), it was when analyzing the sets together ($\beta = -0.08$, $z = -2.32$, $p = 0.02$), suggesting we had been underpowered to detect this effect.

There was no effect of RPE sign on sequence memory (first set: $\beta = -0.03$, $z = -0.42$, $p = 0.68$; replication set: $\beta = -0.03$, $z = -0.46$, $p = 0.65$). We next tested for an interaction between RPE sign and condition to determine whether positive and negative RPEs differentially modulate sequence memory when the RPE indicates a change in reward state versus when it doesn't (i.e., high-versus-low RPE events). We did not find an interaction in either set (first set: $\beta = -0.25$, $z = -1.80$, $p = 0.07$; replication set: $\beta = -0.12$, $z = -1.01$, $p = 0.31$). We moreover did not find that a continuous measure of RPE predicted sequence memory (first set: $\beta = 0.003$, $z = 1.14$, $p = 0.25$; replication set: $\beta = 0.003$, $z = 1.69$, $p = 0.09$). For a discussion of these results, see 'General Discussion'.

Distance memory

We did not find that high RPEs influenced distance judgments in the first dataset ($\beta = -0.01$, $t = -0.52$, $p = 0.61$; Figure 4.3C). To further assess whether the magnitude of the RPE influenced perceived distance, we correlated distance judgments with the reward difference between the pair of items within a pair (which is a proxy for the magnitude of any intervening RPE event, since item values were roughly stable on each side of a high-RPE event). We did not find this measure to predict perceived distance either ($\beta = 0.03$, $t = 0.89$, $p = 0.38$; Figure 4.3E).

In the larger replication dataset, however, we did find two main effects and an interaction between RPE event and presentation distance in modulating distance judgments (Figure 4.3D). Here, perceived distance was higher when the pair included/spanned a high (vs. low) RPE event (RPE: $\beta = 0.11$, $t = 3.12$, $p = 0.002$). These two effects interacted such that the high-RPE effect was strongest for items that were closer together ($\beta = -0.04$, $t = -2.05$, $p = 0.04$), and in particular for pairs that included the high-RPE item itself ('0 between': high RPE_{-1} and high RPE; $\beta = 0.12$, $t = 2.91$, $p = 0.004$, μ-high-RPE = 3.07, μ-low-RPE = 2.90). We also found that the greater the reward difference between the two images, the greater the perceived distance ($\beta = 0.04$, $t = 2.83$, $p = 0.005$; Figure 4.3F). This effect was again largely driven by the pair that included the high RPE event (0 between: $\beta = 0.07$, $t = 2.45$, $p =$

0.01; 1 between: $\beta = 0.05$, $t = 1.85$, $p = 0.06$; 3 between: $\beta = 0.01$, $t = 0.56$, $p = 0.58$).

We did not find that RPE sign predicted distance judgement in the first set ($\beta = 0.05$, $t = 1.30$, $p = 0.20$), while we did find that positive RPEs were associated with increasing perceived distance in the replication set ($\beta = 0.06$, $t = 1.95$, $p = 0.05$). Lastly, we did not find that RPE sign differentially modulated perceived distance for high- versus low-RPE events in either set (first set: $\beta = -0.06$, $t = -0.77$, $p = 0.45$; replication set: $\beta = -0.08$, $t = 1.26$, $p = 0.21$).

Discussion

In Experiment 4 and its replication, we again found that high-RPE events act as event boundaries by interrupting the sequential integration of events into memory, leading to worse sequence memory for events across a high RPE event. Interestingly, and in line with our recognition priming results, there were no differences in sequence memory for the pair that included the high-RPE item itself (i.e., the pair testing the association between the high-RPE$_{-1}$ and high-RPE event), again suggesting that the high-RPE event is associated with its predecessor. Moreover, and in line with a widely reported finding in temporal memory (Yntema & Trask, 1963; Fortin, Agster, & Eichenbaum, 2002; Kesner, Hunsaker, & Ziegler, 2010; Rouhani et al., 2018), we found that greater distance between items improved sequence memory. However this pattern was only present for the low-RPE pairs, providing further evidence that high-RPE events disrupt temporal memory.

Our distance judgement measure yielded mixed results. We only found an effect of high RPEs on perceived distance in the replication dataset: High RPEs led to greater perceived distance, and (relatedly) greater differences in reward value between the two items were associated with greater perceived distance; importantly, these effects were most reliably present for the '0 between' condition, where the pair included the high-RPE event itself. Here, the effects of high RPEs on sequence memory (and recognition latency) showed the opposite pattern from perceived distance: when testing temporal associations between the high-RPE event and its predecessor, evidence of an event boundary is strongest in distance judgment but absent in sequence memory (as well as in recognition memory: Exp. 1 and 2). This qualitative difference suggests a potential dissociation between the mechanisms supporting sequence and distance judgments (Clewett, Gasser, & Davachi, 2019). In this replication set, we also found that positive RPEs led to greater perceived distance, although this effect was not contingent on whether the RPE signaled a change in reward state or not (i.e., high versus low RPE). Nevertheless, as we did not find these distance effects in the first dataset, they require replication.

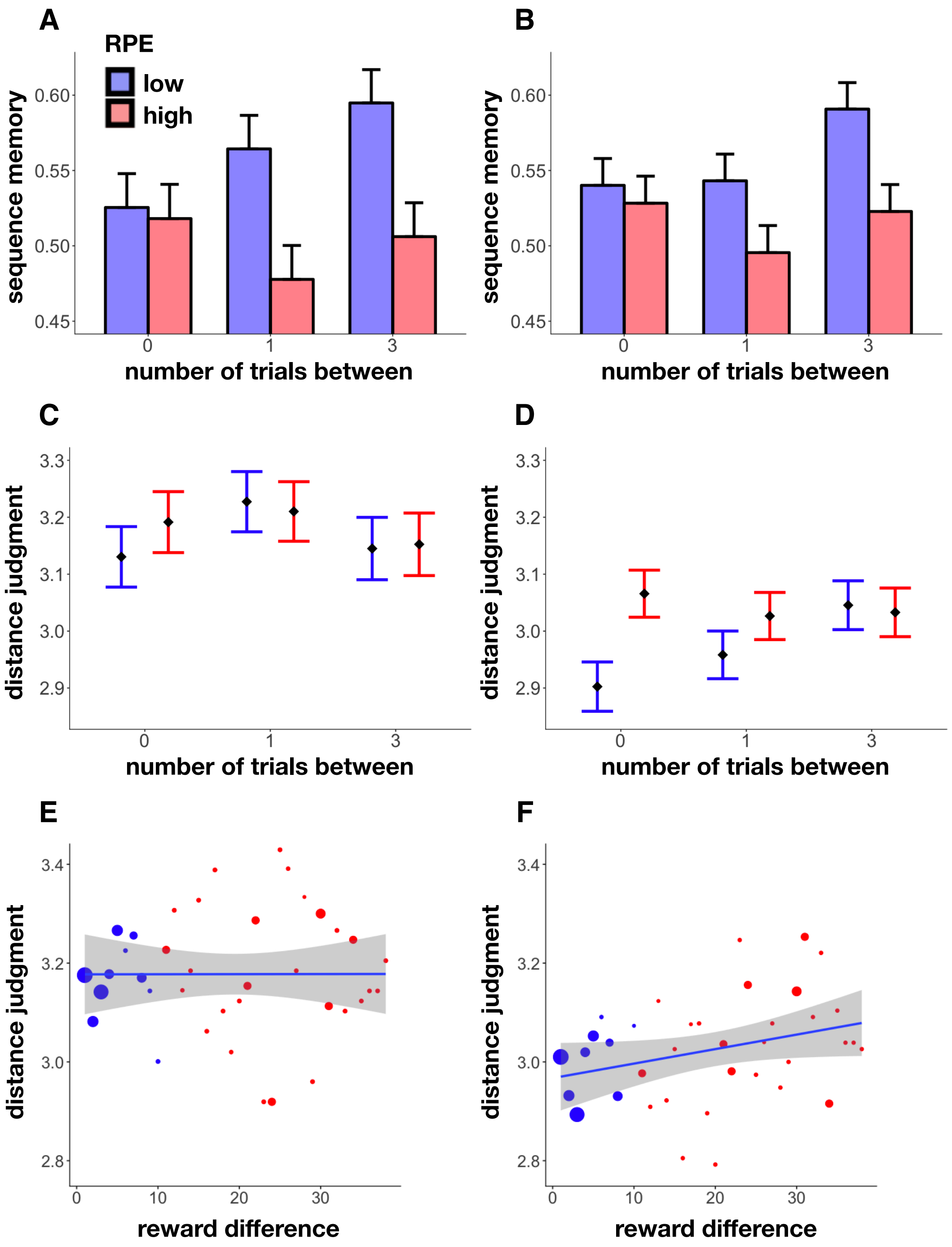


Figure 4.3: Sequence and distance memory results. A-B. Sequence memory in Exp. 4 (A) and its replication (B) as a function of RPE event and presentation distance (number of trials) within scene pairs. Sequence memory for pairs that spanned a high-RPE event was worse; this was driven by pairs that did not include the high-RPE event itself (i.e., pairs that were 1 or 3 trials apart). C-D. Distance judgement in Exp. 4 (C) and its replication (D) as a function of RPE event and presentation distance (number of trials) within scene pairs. High-RPE events were perceived as more distant from each other only in the replication experiment, a result driven by pairs that included the high-RPE event itself (i.e., 0 trials apart). E-F. Distance judgement as a function of the reward difference between scenes in Exp. 4 (E) and its replication (F). In the replication experiment, we found that greater reward difference between scenes, which was a proxy for the magnitude of the RPE event that had occurred between them, led to greater perceived distance. Note that no statistics were run on these averaged values, and they are plotted here for illustration only. Size of the dots reflects the size of that sample. Shaded regions reflect 95% confidence intervals. Error bars represent SEM.

4.5 Computational Model

Overview

To explore potential mechanisms for our findings, we developed a variant of the Context Maintenance and Retrieval model (CMR: Polyn et al., 2009; for other variants, see CMR2: Lohnas, Polyn, & Kahana, 2015, eCMR: Talmi, Lohnas, & Daw, 2019), and tested whether our behavioral results can be explained by a model in which high RPEs induce mnemonic separation between events. In our model, experienced events are temporally linked through a slowly drifting internal 'context', where features of the experienced items update the context representation (Howard & Kahana, 2002). We posit that high RPEs temporarily increase the context drift rate (i.e., the extent to which the high-RPE event updates context), thereby creating a large shift between the context representation of events experienced prior to the high RPE and those after it. We show that this discontinuity can explain our findings of reduced recognition priming.

We used the model to simulate both recognition priming and sequence memory. To simulate recognition priming, we first presented a recognition prime to the model, which triggered an update to the model's context representation. Next, the recognition target was presented to the model. Importantly, activation was allowed to spread back from the context representation (which had been updated by the prime) to the representation of the target; this spreading activation affected the latency with which the target was recognized (for details, see 'Recognition priming' below). For sequence memory, we used a mechanism whereby primacy judgments ("which came first?") were based on which item's context was more distant from the retrieved context (for details, see 'Sequence memory' below).

Representational structure

The model includes two layers, a feature layer (F) and an internal, temporal context layer (C), both of which contain the same number of units. External events (happening at time i) activate a single localist feature in F (f_i), and these activations spread up from F to C (the context layer at time i is denoted as c_i) via a feature-to-context matrix (M^{FC}) that updates context during both the initial encoding phase and the test phase. During retrieval, activations spread back down from C to F via a context-to-feature matrix (M^{CF}) that guides memory search (Figure 4.4). We represent different events as orthogonal unit vectors ('one-hot'). Although the CMR uses an additional 'source layer' to tag explicit contextual shifts (such as different encoding tasks), in our model we did not use this layer to tag different reward states. This is because changes in the reward distribution were latent to the participant (and thus also to the model).

Each associative matrix was made up of an episodic and a semantic component, meaning that M^{CF} comprised a weighted average of episodic (EM^{CF}) and semantic (SM^{CF}) weight matrices, and likewise M^{FC} comprised a weighted average of EM^{FC} and SM^{FC} (we modeled the weights of each matrix separately). As in TCM and CMR, the episodic matrices are updated during encoding to store associations between active feature representations in F and context representations in C. The semantic matrices contain one-to-one connections between a unit in F to its corresponding unit in C (concretely, they are identity matrices).

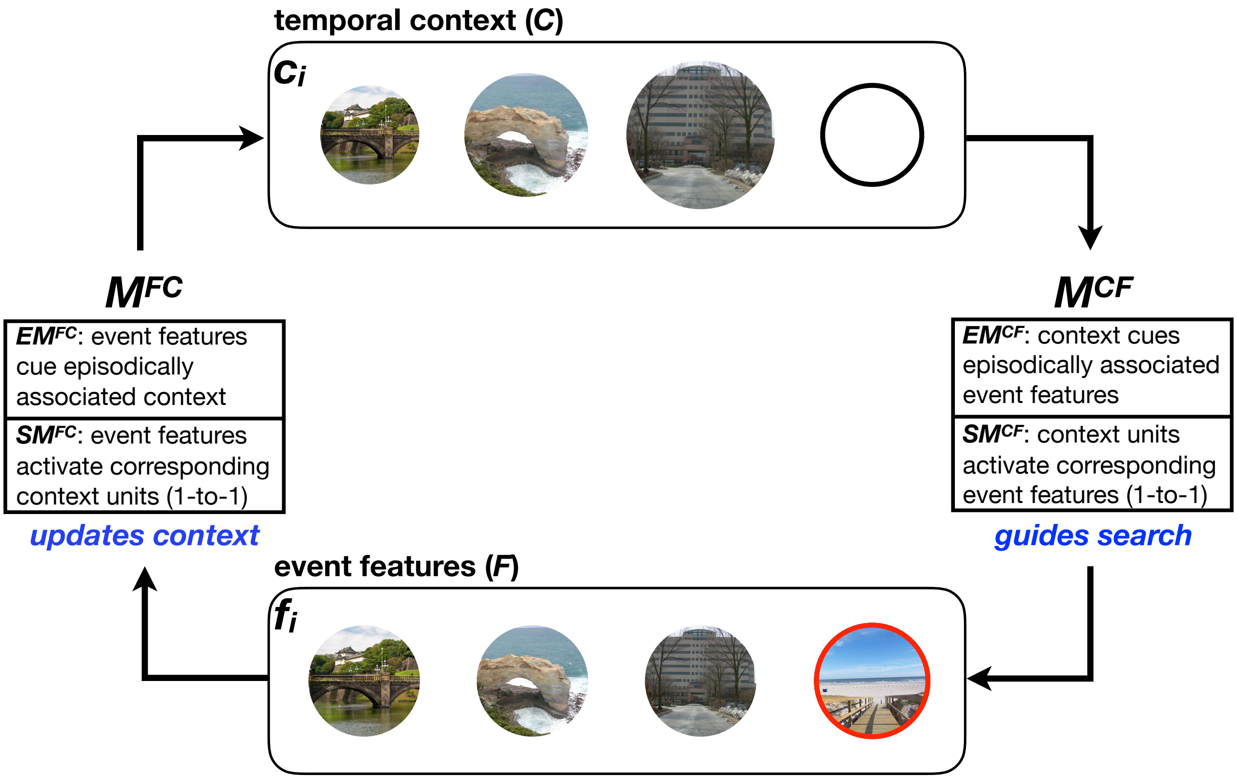


Figure 4.4: Model structure. The model has two layers: a feature layer (*F*) and a temporal context layer (*C*) that interact through two associative matrices: a feature-to-context matrix (M^{FC}) that updates context and a context-to-feature matrix (M^{CF}) that guides search. Each matrix is a composite of an episodic (EM^{CF}, EM^{FC}) and semantic matrix (SM^{CF}, SM^{FC}). The episodic matrices represent the episodic associations formed between *F* and *C* during encoding, whereas the semantic matrices contain one-to-one connections between features in *F* and the corresponding units in *C*. When an event is "experienced" (during encoding) or "remembered" (during retrieval), its corresponding unit f_i is activated in *F*, and activation spreads up to *C* via M^{FC}. Specifically, EM^{FC} updates *C* with contexts that were previously (episodically) linked to f_i ('mental time travel'), and SM^{FC} updates *C* by activating the unit in the context layer that directly corresponds to f_i (e.g., if f_i is the third unit in the feature layer, SM^{FC} activates the third unit in the context layer). During retrieval, activation spreads down from *C* to *F* via M^{CF}. Specifically, EM^{CF} activates units in *F* that were previously (episodically) linked to contexts that match the current state of *C* ('episodic retrieval'), and SM^{CF} activates units in *F* proportionally to how active the corresponding units are in *C* ('direct readout'). Units in *F* then compete for retrieval. The figure depicts the state of the model at time point *i*= 4: The first three items (from left to right) were presented successively on previous trials, and are therefore active in context (more recently experienced items are more active in *C*, as reflected here by the size of the circles); the fourth item (outlined in red) is being presented in the feature layer. This feature-layer representation of the fourth item will be episodically associated with the context shown here; on the next time step it will be used to update the state of *C* (via M^{FC}) and the cycle will begin again.

Updating temporal context and associative matrices during reward learning

Prior to the reward learning phase, C and the episodic associative matrices (EM^{CF} and EM^{FC}) are initialized to zero. When an item is activated in F during the reward learning phase, the activation spreads up from F to C via M^{FC} where the input to C is calculated as follows:

$$c^{IN} = M^{FC} f_i. \tag{4.1}$$

The vector c^{IN} is then normalized to be of unit length, and then context is updated as follows (as in TCM and CMR):

$$c_i = \rho_i c_{i-1} + \beta c^{IN}, \tag{4.2}$$

$$\rho_i = \sqrt{1 + \beta^2[(c_{i-1} \cdot c^{IN})^2 - 1]} - \beta(c_{i-1} \cdot c^{IN}). \tag{4.3}$$

Here, β defines the degree to which the active feature causes the context to 'drift' – the larger the value of β, the more the active feature will be inserted into the context, crowding out other active events in C. We allowed for two distinct drift values, β, the standard drift (implemented for low-RPE events), and d, a higher level of drift for high-RPE events. This approach (i.e., increased drift in response to high-RPE events) is in line with how contextual disruptions due to salient changes have been previously modeled (Horner et al., 2016; Polyn et al., 2009; Siefke, Smith, & Sederberg, 2019). We moreover use d for the first item presented to the network as a way of capturing classic primacy effects in memory (i.e., the higher probability of retrieving the first item in a sequence; see 'Model calibration' for further discussion of how primacy is modeled here, compared to how it is usually modeled in CMR).

The two episodic associative matrices are updated through Hebbian outer-product associative learning. α represents the learning rate for that update:

$$\Delta EM^{FC} = \alpha^{FC} c_i f_i^{\ T}, \tag{4.4}$$

$$\Delta EM^{CF} = \alpha^{CF} f_i c_i^{\ T}. \tag{4.5}$$

Importantly, in our version of the model, on each time step, the following order-of-operations applies: First, the feature vector is updated based on the current event; next, the episodic matrices are updated; and finally the context vector is updated. The consequence of this order-of-operations is each event is inserted into the *following* event's episodic context (but not its own episodic context). For example, at the end of the fourth time step, the fourth item will be inserted into the context layer; at the start of time step 5, the fifth item's feature-layer representation will be activated, at which point it will be episodically associated with the current state of the context layer (where the *fourth* item's context-layer representation is active). Next, the fifth item's context-layer representation is activated, and the cycle begins again. We also simulated our results with a version of the model where context is updated before the episodic matrices (like CMR), and were not able to capture our effects (Figure 4.5).

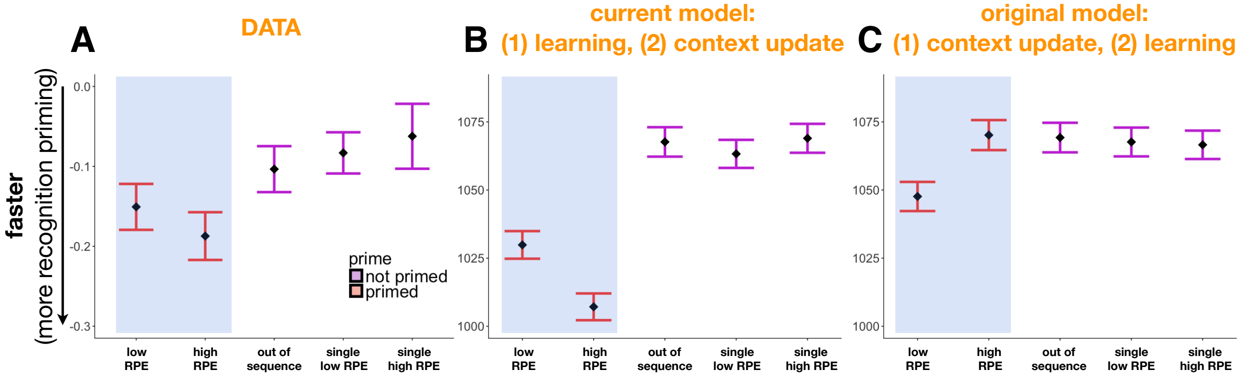


Figure 4.5: Behavioral results (A), simulation results with the order of operations from the current model (B), and simulation results with the order of operations from the original Polyn et al. (2009) model (C). Note that we use the same model parameters (chosen based on our order of operations) in both cases. The shaded blue areas emphasize the critical difference in the predictions of the two models: The original order of operations (C) creates a representational discontinuity between the high-RPE event and its predecessor, because the high-RPE event drifts into its own context during encoding; this has the effect of disrupting recognition priming for the high-RPE event when primed by its predecessor. By contrast, our model (and the actual data) show intact recognition priming in this condition.

Simulating free recall

Although we did not collect our own free recall data, we calibrated the model by running free-recall simulations, using the following procedure. First, after the learning trials, we simulated the intervening time period before the memory test by presenting 15 randomly-generated 'distractor' events. This allowed active features in C to substantially drift from the encoding period, thereby capturing the putative drift occurring between the end of the learning phase and the start of the test phase. These distractor events did not compete during retrieval.

The associative matrices at recall were each calculated as a weighted average of their episodic and semantic components:

$$M^{FC} = \gamma^{FC} SM^{FC} + (1 - \gamma^{FC}) EM^{FC}, \tag{4.6}$$

$$M^{CF} = \gamma^{CF} SM^{CF} + (1 - \gamma^{CF}) EM^{CF}, \tag{4.7}$$

As in CMR and TCM-A (Sederberg, Howard, & Kahana, 2008), the recall period was governed by a leaky, competitive accumulation process where experienced events accumulated activation until one passed a threshold and 'won' the competition (Usher & McClelland, 2001). The following calculates the input to the accumulators:

$$f^{IN} = M^{CF} c_i. \tag{4.8}$$

Which then guides the below competition dynamics:

$$\begin{gathered} x_s = (1 - \tau\kappa - \tau\lambda N)x_{s-1} + \tau f^{IN} + \varepsilon, \\ x_s \rightarrow max(x_s, 0). \end{gathered} \tag{4.9}$$

Here, x is a vector with units corresponding to each element in the feature layer (f^{IN}), and s indexes the step in the accumulation process (units are initialized to zero, and cannot take on negative values, second line of Eq. 4.9). The parameters governing the competition are τ, the time constant determining the rate of accumulation, κ, the decay rate for active items, and λ, the lateral inhibition parameter which scales the strength of inhibitory matrix, N; ε adds gaussian noise to the decision process (drawn from a random normal distribution with mean zero and standard deviation η). This accumulation process proceeded until one of the elements passed a threshold of 1, at which point the winning item's feature was reinstated in F, and its encoding context was reactivated, using Eq. 4.1.

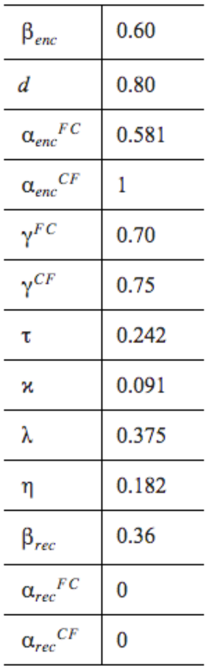
The reactivated context was then used to update the current context vector following Eq. 4.2. Subsequently, f^{IN} was updated and the accumulation process restarted with $x_1 = 0$. Previously retrieved items were allowed to continue competing in the accumulation process, but were prevented from passing the retrieval threshold.

Model calibration

Before simulating our experiments, we determined which parameter values to use by identifying combinations that replicate canonical findings in free recall tasks; namely, the higher probability of recalling the first item ('primacy') and the last item ('recency') in a given context, along with contiguity effects (increased likelihood of recalling items that were studied close together in time, with a bias towards forward transitions (Howard & Kahana, 2002). We identified these parameters by feeding our network distinct events (orthogonal one-hot vectors) and running network simulations for all value combinations of the following four parameters (ranging from 0-1, in increments of 0.05; 100 simulations for each combination): (1) d, context drift for primacy events (and for high-RPE events, in the recognition simulations presented later); (2) β, context drift for non-primacy events (and for low-RPE events in the recognition simulations); (3) γ^{CF}, the relative weight assigned to the semantic vs. episodic components in M^{CF}; and (4) γ^{FC}, the relative weight assigned to the semantic vs. episodic components in M^{FC}. All other parameter values were taken from Polyn et al., 2009; see Table 4.1. We generated serial position curves and conditional response probability curves for each run, and filtered the parameter values based on whether they generated characteristic features of these recall curves (Figure 4.6). Specifically, in the serial position curves, the parameter values we chose generated primacy (higher recall of the first item relative to the subsequent one) and recency effects (higher recall of the last item relative to the preceding one). When simulating contiguity effects, we looked for parameter values that resulted in greater sequential recall of events that were neighboring during encoding, with an increased likelihood of forward recall (thereby matching the pattern that is typically observed in free recall (Howard & Kahana, 2002).

We found that recency and contiguity effects were obtained across a fairly wide range of parameters in the model (as has been shown in previous work with TCM and CMR (Howard & Kahana, 2002; Polyn et al., 2009). Primacy effects were obtained across a more narrow range of parameters. Specifically, to obtain primacy effects we needed to have a relatively high drift rate for primacy items (d) compared to the drift rate for non-primacy items (β), as well as a strong contribution of the semantic matrix to both M^{CF} and M^{FC} (i.e., high values of γ^{CF} and γ^{FC}). This configuration of parameters allowed primacy effects to arise in the following manner: When the primacy item is present, it is strongly inserted into context, due to the high value of context drift (d) that we assigned to primacy items, and the high contribution of the semantic matrix to M^{FC}. Because the primacy item is strongly inserted into context, it is still present in context (i.e., its unit's activation has not fully decayed away) at the time of test. Because of the strong contribution of the semantic matrix to M^{FC} (which supports 'direct read-out' of active items in context back into the feature layer), the fact that the primacy item is still active in context leads to increased activation of that item back in the feature layer (via the aforementioned 'direct readout' mechanism; see 'Simulation results' for how these matrices interact during our simulations). Note that this way of modeling primacy is different from how primacy is handled in CMR (Polyn et al., 2009) – primacy items are assigned a higher learning rate (for forming episodic context-to-feature associations) but the drift rate is the same for primacy and non-primacy items. A key goal of our modeling exercise was to assess if we could model our own experimental results and also classic recall effects (e.g., primacy) only through drift manipulations and not through learning rate manipulations; we return to this point in 'Summary of computational model' in the 'General Discussion' below.

As a result of these initial simulations, we selected the following parameter values: $d = 0.8$; $\beta = 0.6$; $\gamma^{CF} = 0.75$; $\gamma^{FC} = 0.70$. We subsequently ran the recognition priming and sequence memory procedure detailed below using these parameters. For recognition priming, we ran 10,000 simulations for each condition, and for sequence memory, we ran a single simulation for each condition since dynamics during encoding are deterministic.



β_{enc}	0.60
d	0.80
α_{enc}^{FC}	0.581
α_{enc}^{CF}	1
γ^{FC}	0.70
γ^{CF}	0.75
τ	0.242
$\varkappa$	0.091
λ	0.375
η	0.182
β_{rec}	0.36
α_{rec}^{FC}	0
α_{rec}^{CF}	0

Table 4.1: Parameters used in model. β_{enc}, d, γ^{FC}, and γ^{CF} were determined based on a parameter search for values that best captured primacy and recency effects in serial position curves, as well as the signature characteristics of conditional response probability curves in free recall. All other parameters were taken from Polyn et al., 2009; '*enc*' refers to the value used during encoding and '*rec*' during recall.

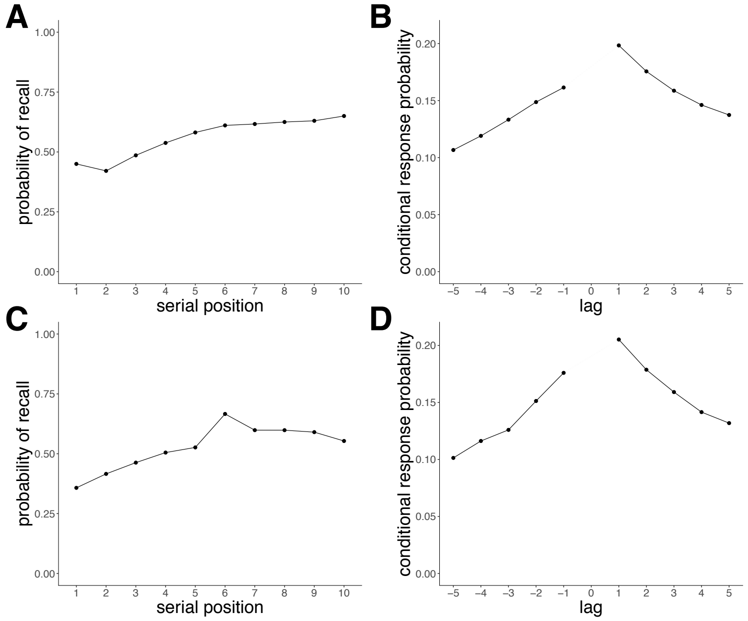


Figure 4.6: Free recall simulations for a sequence of items without a high-RPE event (A-B) and with a high-RPE event at item 6 (C-D); 10,000 simulations each. A. Serial position curve for a low-RPE sequence retrieved after another 15 ‘distractor’ items. There is a higher probability of retrieving the first item than the second item (primacy) as well as for retrieving more recent items (recency). Note the curve is much flatter than typical immediate free recall results due to the substantial number of distractor events between encoding and retrieval. B. Conditional response probability (CRP) curve for a low-RPE sequence retrieved after 15 distractor items. The model captured the main characteristic of CRP curves: Retrieval of an item led to a greater likelihood of recalling items that were nearby during encoding, and in particular for items that were encoded after the recalled item (forward asymmetry). C. Serial position curve for a high-RPE sequence retrieved after 15 distractor items. The high-RPE item is the sixth in the sequence. The higher drift associated with the high-RPE event led to a greater likelihood of retrieving that item and diminished the primacy effect. D. Conditional response probability (CRP) curve for the high-RPE sequence showing similar characteristics to those of the low-RPE sequence.

Recognition priming

To simulate our recognition priming results, we used the following procedure: After the initial learning phase and presentation of filler items, we presented a ‘recognition prime’ (a low- or high-RPE$_{-1}$ event) to the network by activating the ‘one-hot’ feature vector that represents that event. After the prime’s representation was activated in F, activation was allowed to spread up from F to C via M^{FC}. The EM^{FC} component of M^{FC} updates the context vector with the prime’s episodic context (i.e., the context linked to the prime at encoding; this is the process commonly referred to as ‘mental time travel’, since it makes the context at test resemble the context when the prime was studied (Kragel, Morton, & Polyn, 2015; Tulving, 1984)). The SM^{FC} component of M^{FC} allows for the prime itself to be inserted into C (see ‘Simulation results’ and the subsequent discussion for more description on how these matrices interact during retrieval).

Note that prior studies have found that recognition priming is only obtained when the prime is successfully recollected at test (Schwartz et al., 2005). Our allowing activation to spread from the prime's feature-layer representation to C via EM^{FC} corresponds to an assumption that the prime was (itself) successfully recollected; this assumption is justified because – in the priming data that we set out to model – we only analyzed trials where the prime was successfully remembered (so the assumptions of the model match the structure of our analysis).

After context was updated by the prime, the recognition trial was simulated. Here, activation was allowed to spread down from C to F via M^{CF}. EM^{CF} modulates item activation as a function of the match between each item's episodic context and the current context, and SM^{CF} provides a 'direct readout' of activations from C to F (e.g., if the fifth unit in C is active, activity spreads directly down to the fifth unit in F). We then allowed the competition dynamics to unfold. To simulate the fact that the recognition target is presented perceptually, we boosted the activation of the target event by in F by 0.75 at the start of the competition; this had the effect of ensuring that the target event would be the winner of the competition, but still allowed for variance in recognition latency. We extracted recognition latencies for the target item and compared them with the empirical recognition data.

We tested target items matched to our experimental conditions, and ran simulations for each condition separately (Figure 4.7). For the simulation of Experiment 2, the 'low RPE' target was the low-RPE event that had been studied directly after the prime and the 'high RPE' target was the high-RPE event that had been studied directly after the prime thus testing the link between the high-RPE event and its predecessor. In this simulation, we further tested the associative links between the high-RPE event and its successor (the high-RPE$_{+1}$), which we did not behaviorally test; here, the prime was the high-RPE event and the target was the high-RPE$_{+1}$ event. For the simulation of Experiment 3, the low-RPE$_{+1}$ target was a low-RPE event that had been studied two events after the prime, and the high-RPE$_{+1}$ target was a low-RPE event that had been studied two events after the prime (with the high-RPE event having occurred between the prime and the target). The 'out of sequence' target was always an event that had been studied before the prime (3 trials apart). For conditions where there was no prime ('single high-RPE' and 'single low RPE'), we did not present a prime to the model prior to simulating target recognition – in this case, the state of C the start of target recognition only reflected the effects of the reward learning phase and the distractor items (but not the prime); otherwise, the procedure was the same as in primed trials.

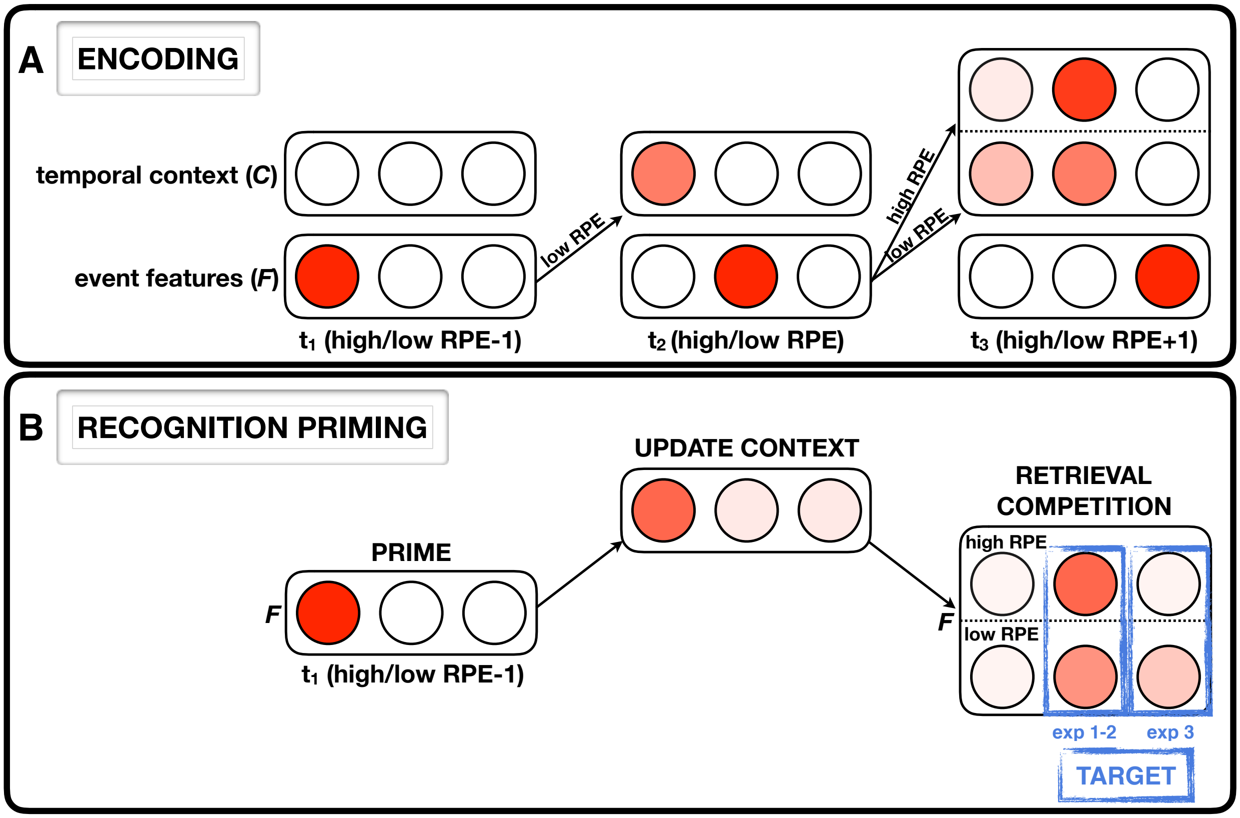


Figure 4.7: Illustration of how high- versus low-RPE events are encoded and retrieved by the model. A. Encoding. The high-RPE event enters the next event's context with increased drift, leading to greater activation of the high-RPE event at the expense of the activation of the high-RPE$_{-1}$ item in context. Learning of context-feature associations in the model is based on co-activity of context and feature units (Hebbian learning); because the high-RPE$_{-1}$ item is less active in context, it becomes less strongly associated with the high-RPE$_{+1}$ item in the feature layer. B. Recognition priming. (1) The prime (high/low RPE$_{-1}$ event) is retrieved by the network, (2) Activation spreads up from F to C via M^{FC}, leading to higher activation of the prime in context, (3) Activation then spreads down from C to F via the M^{CF}, leading to activation of events that contained the prime in their context, (4) The prime strongly cues both the high/low-RPE targets (Exp. 1 and 2). However, when the target item is the high-RPE$_{+1}$ event (Exp. 3), that item receives less activation because of the weaker association between the high-RPE$_{-1}$ item (in context) and the high-RPE$_{+1}$ item (in the feature layer), as mentioned above.

Sequence memory

We aimed to capture the two characteristic results of our sequence memory paradigm: (1) worse sequence memory for items that span a high-RPE event, and (2) better sequence memory with greater distance between items in the low-RPE condition (but not the high-RPE condition).

Our simulation was based on 'distance theories' of temporal order memory (Friedman, 1993, 2004), which posit that judgments of recency are a function of the similarity of the context associated with an item at encoding and the context that is active at test (such that greater similarity leads to judgments of greater recency (Hintzman, 2002)). In our task, we did not ask for a judgment of recency, but instead asked for a judgment of primacy. To simulate these judgments, we first retrieved the context vectors associated (at encoding) with each tested item, and then correlated each of these retrieved vectors with the context vector active at test. We next took the difference of these correlations within each tested pair, and used this as a measure of distance between items. We then put this correlation difference through a sigmoid function whereby larger differences increased the likelihood of a primacy judgment for the more contextually distant item.

Simulation results

During the initial encoding (i.e., reward learning) phase, our use of a higher drift rate for high-RPE events created a discontinuity in the mental contexts associated with events that occurred before the high RPE event versus those that occurred after it. We tested how this representational 'event boundary' affected recognition priming in simulations of Experiments 2 and 3. Experiment 2 (Figure 4.8A-B) tested pairs of events that were consecutively-encoded during the reward-learning phase – call these events n and $n+1$ (referring to their adjacent positions during learning). For some pairs, event $n+1$ was a high RPE event ('high RPE'), and for other pairs, event $n+1$ was a low RPE event ('low RPE'). As noted in 7.4, the model is set up such that (during reward learning) each item becomes part of the next item's episodic context (i.e., item n is strongly active in the context layer when item $n+1$ is activated in the feature layer; see Figure 4.7). At test, when item n is presented as a prime (by activating its representation in the feature layer), activation spreads up to item n's representation in the context layer (via the influence of SM^{FC}). Next, activation is allowed to spread back down to the feature layer via M^{CF}. Here, the influence of EM^{CF} is crucial – the effect of this matrix is that items whose context at study matched the current context are activated in the feature layer. Crucially, because item n was part of item $n+1$'s context at study, the effect of EM^{CF} in this situation is to allow activation to spread from the 'item n' unit in the context layer to the 'item $n+1$' unit in the feature layer. This spreading activation allows the 'item $n+1$' unit to cross threshold sooner when item $n+1$ is presented as a recognition target, thereby giving rise to the recognition priming effect.

This priming effect is present in the model for both high-RPE primed targets and low-RPE primed targets, but it is larger in magnitude for high-RPE targets than low-RPE targets (μ-high-RPE-primed = 1007.14 ms, μ-low-RPE-primed = 1029.86 ms, μ-out-of-sequence = 1067.65). Moreover, there was an interaction in the retrieval of high and low-RPE targets that were primed versus those that were not, indicating that priming led to the faster retrieval of the high-RPE target relative to the low-RPE target (μ-high-RPE-single = 1068.98 ms, μ-low-RPE-single = 1063.26 ms). The difference in priming effects (in the model) between high-RPE and low-RPE targets is caused by the influence of SM^{CF} at retrieval. In addition to the effects of EM^{CF} (described above), SM^{CF} provides a 'direct read-out' of which items are active in the context layer. Because of the greater drift associated with high-RPE items, high-RPE (vs. low-RPE) items end up being more strongly active in context (even at the time of test). This extra activation in context translates (via the influence of SM^{CF}) into greater activation of the high-RPE target in the feature layer, which further speeds recognition for high-RPE items, boosting the level of recognition priming.

In addition to the strong link between the high-RPE event and its predecessor, we found the high-RPE event to be similarly linked to its successor. In fact, when primed by the high-RPE event, the high-RPE$_{+1}$ event was retrieved faster than when the high-RPE event was primed by its predecessor (μ-high-RPE$_{+1}$-primed = 993.18). This is because the high-RPE item itself gets strongly inserted into the high-RPE$_{+1}$ item's context during encoding, and then subsequently during retrieval, priming the network with the high-RPE event leads to strong forward retrieval of the high-RPE$_{+1}$ event. Although we did not test this association behaviorally, our model simulation therefore suggests that the high-RPE event is linked to both its predecessor and its successor.

In Experiment 3 (see Figure 4.8C-D), primed target items were studied two items after the prime during the learning phase (i.e., with one event in between); sometimes the event interposed between prime and target during learning was a high-RPE event, and sometimes it was a low-RPE event. For the purpose of explaining what happens in the model on these trials, call the prime item *n*-1, the interposed item *n*, and the target item *n*+1. First, consider the condition where the interposed item was a low-RPE event. In this case, during learning, item *n*-1 (the prime) is still strongly active in context when item *n*+1 (the target) is studied, so the prime's representation in context gets linked to the target's representation in the feature layer. Because of this link, the usual mechanisms of recognition priming (as described in the preceding paragraph) still apply. Next, consider the condition where the interposed item was a high-RPE event. Because of the higher drift rate for high-RPE items, the effect of (strongly) inserting high-RPE item *n* into context is to 'push out' the representation of item *n*-1 from the context layer. Because item *n*-1 (the prime) is no longer strongly active in context when item *n*+1 (the target) is studied, the crucial episodic link between the prime (in context) and the target (in the feature layer) is not formed, eliminating the recognition priming effect. Finally, there was an interaction in the retrieval of primed high and low-RPE items between experiments, such that priming of the high-RPE event itself (Exp. 2) is enhanced whereas priming of the high-RPE$_{+1}$ event (Exp. 3) is interrupted relative to the low-RPE items.

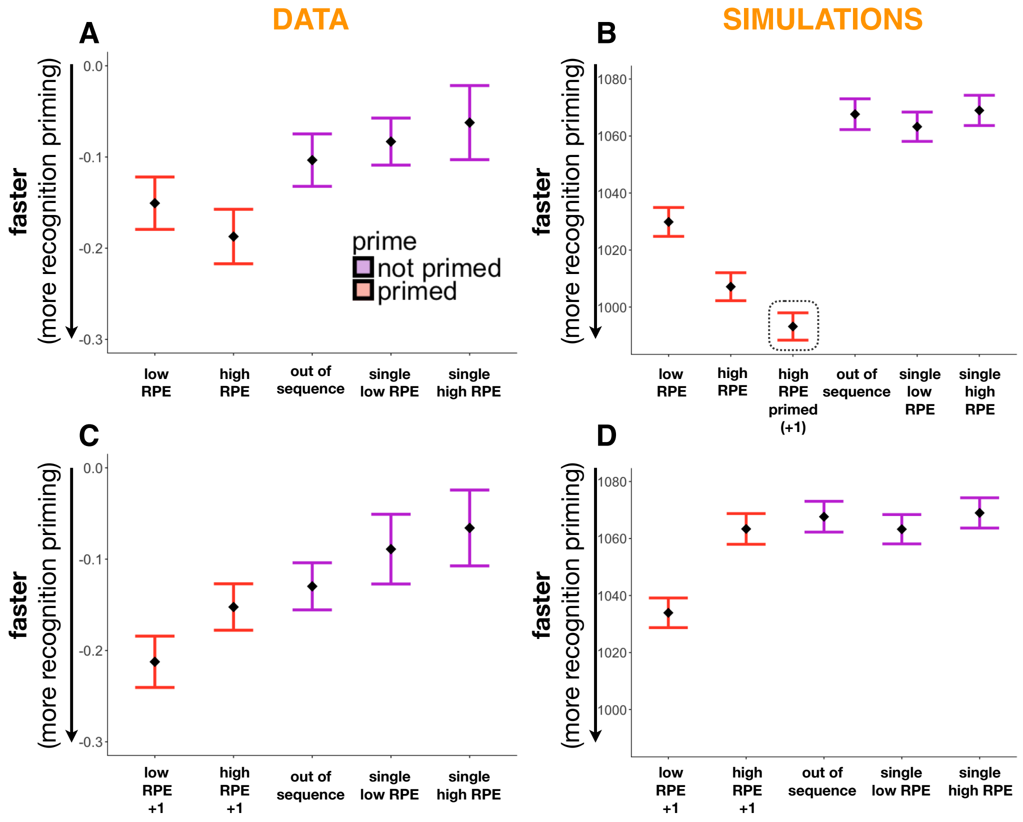


Figure 4.8: Recognition-priming simulation along with behavioral results. A-B. Recognition latencies as a function of item condition in Exp. 2 (A) compared to model simulations (B). High- and low-RPE targets are retrieved faster than the out-of-sequence targets. In the simulations there is, moreover, an interaction between priming condition and RPE, such that primed high-RPE targets are retrieved faster than primed low-RPE targets, but this difference was not observed for single (unprimed) items. The data point in the dotted outline represents recognition priming for the high-RPE$_{+1}$ item (target) when primed by the high-RPE event (prime): although we did not test this behaviorally, the model predicts a strong association between the high-RPE event and its successor. C-D. Recognition latencies as a function of item condition in Exp. 3 (C) compared to model simulations (D). In both the data and the simulations, the high-RPE$_{+1}$ target no longer shows recognition priming (i.e., it is no longer retrieved faster than the out-of-sequence target) but the low-RPE$_{+1}$ target shows robust recognition priming.

In our sequence memory simulation, primacy judgments were based on which item's context was more distant from the retrieval context. Sequence memory for low-RPE pairs therefore improved the further the items had been from each other at encoding (μ-low-RPE-0 = 0.49, μ-low-RPE$_{-1}$ = 0.51, μ-low-RPE-3 = 0.54; Figure 4.9). However, the simulation incorrectly predicted that sequence memory for high-RPE pairs improves with increasing distance between items at encoding. Moreover, it incorrectly predicted better sequence memory for items spanning a high (compared to low) RPE event. This was because - in our model - high RPEs induced a strong shift in context; this context shift between the first and second items selectively increased the contextual distance between the first item and the test context, thereby making the model more (instead of less) accurate at choosing which item came first.

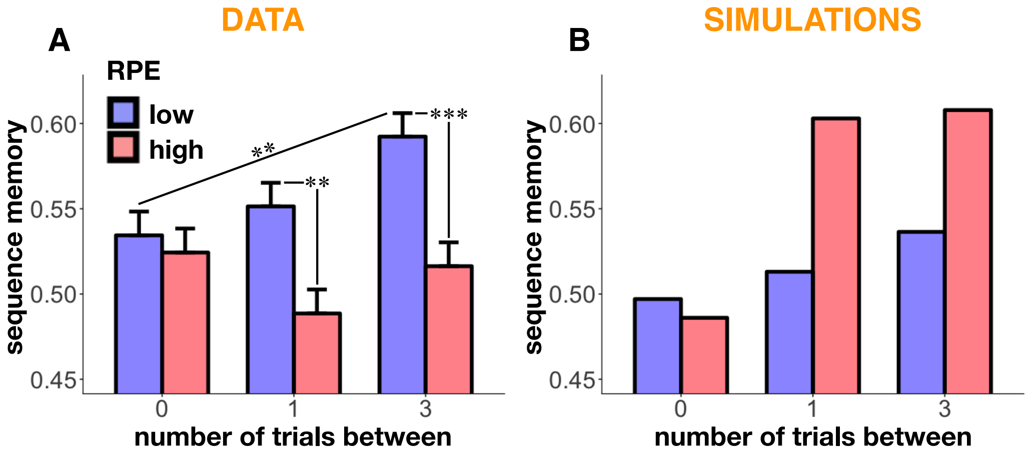


Figure 4.9: Sequence-memory simulations along with behavioral results. A. Sequence memory as a function of RPE event and presentation distance (number of trials) within scene pairs in Exp. 4 and its replication (results averaged across both). Sequence memory was impaired for items spanning a high-RPE event (i.e., 1 and 3-trials-between), but there was no impairment in sequence memory for the high-RPE event and its direct predecessor (i.e., 0-trials-between). Also, sequence memory improved with increasing distance between items in the low-RPE condition but not the high-RPE condition. B. The simulation captured the effect that sequence memory improved with increasing distance between items in the low-RPE condition, but incorrectly predicted that sequence memory would improve with increasing distance between items in the high-RPE condition, and also incorrectly predicted better sequence memory for items spanning a high (compared to low) RPE event. Error bars represent SEM.

Discussion of simulation results

Our model, with parameters chosen to generate canonical free recall dynamics, was able to capture the signature effects of our recognition and sequence memory tasks. In our simulation of Experiment 2, we found that feeding the network recognition primes led to the faster retrieval of target items that had come directly after the primes during the initial reward learning phase (i.e., the low- and high-RPE targets) as compared to targets that were out of sequence or were not primed (single items). Recognition priming was especially strong for high-RPE items, whose higher activation in *C* led to faster retrieval times as compared to low-RPE targets. Although we did not observe significantly faster retrieval times for high-RPE versus low-RPE events in Experiments 1 and 2, the simulation results suggest that the numerical difference in their latencies may reflect an actual effect, which may reach significance with sufficient power. Moreover, the model predicted not only an association between the high-RPE event and its predecessor, but also one between the high-RPE event and its successor.

In our simulation of Experiment 3, the prime and the target always had one event (either high-

RPE or low-RPE) between them. Our model captured the lack of recognition priming in the high-RPE condition by creating a contextual discontinuity after the high-RPE item, thereby 'breaking' the contextual link between the prime and the target. The same mechanism allowed us to simulate the key result from Experiment 4: impaired sequence judgments when the to-be-judged items surrounded a high-RPE event during learning.

In introducing the simulations, we identified four parameters of interest, namely the drift rates for high-RPE and low-RPE events at encoding and the episodic and semantic proportions of the associative matrices. The effects of drift rate on model results ended up being fairly straightforward: d (the high-RPE drift rate) had to be larger than β (the low-RPE drift rate) to create the aforementioned contextual 'gap' after high-RPE items, which is how we explain impaired recognition priming in the high-RPE condition of Experiment 3, and impaired sequence memory in the high-RPE 1-trial-between and 3-trials-between conditions of Experiment 4.

The effects of γ_{FC} and γ_{CF} (episodic/semantic balance in the associative matrices) ended up being more complex. As discussed above, recognition priming depends on the semantic component of M^{FC} and the episodic component of M^{CF}: The prime is loaded into context via SM^{FC}, and then it cues the target via EM^{CF} (since the prime was part of the target's episodic context during learning). Note that this is the same basic sequence of events that accounts for the forward bias in contiguity effects in free recall. The only difference is that, in free recall, the just-recalled item plays the role of the prime: the just-recalled item is loaded into context via SM^{FC} and cues recall of the following item via EM^{CF} (Howard & Kahana, 2002). Thus, to explain recognition priming effects (and forward contiguity effects in free recall), we need to ensure a substantial contribution of SM^{FC} and EM^{CF}.

However, it would be unwise to fully "tilt" M^{FC} towards semantic memory and M^{CF} towards episodic memory. The episodic component of M^{FC} is also important: As noted earlier, this component is what gives rise to 'mental time travel' effects in free recall – in particular, backward transitions in free recall (i.e., recalling items that were studied *before* the item that was just recalled) are thought to result from a sequence where recalling an item reinstates that item's context via EM^{FC}, which then biases recall towards nearby items symmetrically in the backward and forward directions (Howard & Kahana, 2002). The semantic component of M^{CF} is also important: As described in 'Model calibration', our model uses this 'direct readout' component to explain primacy effects in free recall – the primacy item is (strongly) inserted into context via SM^{FC} and then is directly read out from context at test via SM^{CF}. As an aside, this same mechanism that gives rise to primacy would also predict increased free recall of high RPE items (which, like primacy items, are assigned a higher-than-usual drift rate); we have not yet run an experiment to test this prediction in our paradigm.

To summarize the above: Both the episodic and semantic components of both M^{CF} and M^{FC} are important for explaining various effects (either effects in our data or classic regularities in free recall). As such, the greatest challenge in parameterizing the model was finding the right balance between the episodic and semantic components for each matrix. The fact that we found a set of parameters that works well for simulating our results (without impeding our ability to simulate primacy/recency/contiguity in free recall) serves as an existence proof that these factors *can* be suitably balanced.

In our simulation of sequence memory, we sought to explain the following two effects: (1) impaired sequence memory for items that spanned a high-RPE event, and (2) better sequencing of low-RPE pairs (but not high-RPE pairs) the further the items had been from each other at encoding. We implemented a mechanism whereby primacy judgments were based on which item's context was more distant from the retrieval context, in line with distance theories of temporal ordering (Friedman, 1993, 2004). This simulation correctly predicted the better sequencing of low-RPE items that had been further apart during encoding, but incorrectly predicted better, instead of worse, sequence memory for items that spanned a high (compared to low) RPE event. This finding highlights that contextual distance theories can explain sequence memory performance for items within the same latent context, but cannot explain it for items across latent contexts (i.e., across a high-RPE event), suggesting separate mechanisms for organizing events within and across a latent context (DuBrow & Davachi, 2016; Ezzyat & Davachi, 2014).

What kind of model could correctly explain the full pattern of sequence effects we observed? An alternative account of sequence memory is that it relies on explicitly reconstructing the chain of events involving the two queried items. Here, the sequence judgment could be based on the difference in reconstructed order between the two items: P(A before B) would be proportional to the reconstructed serial position of B minus the reconstructed serial position of A. In this kind of model, event boundaries induced by high RPEs should disrupt temporal order memory by making it harder to reconstruct the chain of events (i.e., they create a break in the associative 'chaining' of events across a high-RPE event, Friedman, 1993). Furthermore, the model should also show greater accuracy with increasing distance between the items. While CMR does not include this kind of explicit reconstruction mechanism, other recently developed models, like the Structured Event Memory (SEM) model (Franklin, Norman, Ranganath, Zacks, & Gershman, 2020), do incorporate this mechanism. In future work, we plan to explore how well SEM can account for these sequence memory results.

Lastly, although we found RPEs to modulate distance memory in the replication set of Experiment 4, we did not simulate distance memory results. This was because we did not find consistent results between this experiment and its replication, and because we found a pattern of RPE-modulated effects for distance memory that was opposite to our recognition priming and sequence memory results. For distance memory, we found the greatest difference between the high and low-RPE conditions for the pairs that included the RPE event and its predecessor; by contrast, for recognition priming and sequence memory, effects of high vs. low RPE were largest for pairs that spanned the high-RPE event, and nonsignificant for the pairs that included the RPE event and its predecessor. For this reason, we speculate that distance memory may be supported by a different process than the one we have outlined.

4.6 GENERAL DISCUSSION

Summary of behavioral results

In a passive-viewing, Pavlovian reward learning task, we found that large reward prediction errors (RPEs) enhance memory for that event, demonstrate intact links with preceding events, yet create event boundaries, thereby chunking rewarding experiences into discrete states in memory. Like other types of event boundaries, high RPEs enhance recognition for the event associated with the prediction error, while interrupting memory of the sequence of events across the boundary itself. Specifically, we showed that high-RPE items demonstrate recognition priming, i.e, faster recognition of those items when primed by the previous item, indicating intact associative links with preceding events during encoding (i.e., between high-RPE$_{-1}$ and high-RPE events: Exp. 1-2). However, we found diminished recognition priming for events surrounding the high-RPE item (i.e., between high-RPE$_{-1}$ and high-RPE$_{+1}$ events: Exp. 3) providing evidence of an RPE-modulated event boundary. Moreover, we found that temporal memory, and in particular sequence memory, was worse for pairs that spanned a high RPE versus those that did not (Exp. 4). Interestingly, and analogous to our recognition priming results, this worse overall sequence memory was seen for pairs that excluded the high-RPE event itself, whereas we did not find impaired sequence memory for the pairs that included the high-RPE event and its predecessor.

Summary of computational model

To illustrate and better understand the effects of event boundaries on memory in our experiments, we developed a computational model, a variant of the CMR model (Polyn et al., 2009). To explain the effects of RPEs on memory, our model posits that large RPEs increase the drift rate of contextual information, effectively flushing out previous events and adding the current event into the drifting context.

We simulated recognition priming in the model and analyzed simulated recognition latencies; we also simulated the accuracy of sequential memory judgments. Using the mechanism described above (increased drift in response to large RPEs), we were able to explain our four most important experimental findings: 1) there was recognition priming for pairs of items that were presented sequentially at encoding, regardless of the size of the RPE associated with the target item; 2) when testing for priming of events that were separated by one event during encoding, recognition priming was disrupted if the intervening event triggered a high RPE; 3) when testing sequence memory for pairs of items that were presented adjacently at encoding, having one of those items be a high-RPE event did not impair sequence memory; 4) when testing sequence memory for pairs of items that were presented either one or three items apart at encoding, sequence memory was worse if that gap included a high-RPE event.

These simulation results illustrate the sufficiency of our drift-rate manipulation for explaining the effects of high (vs. low) RPE in the studies reported here. However, this demonstration of sufficiency does not rule out the possibility that RPEs can affect declarative memory in other ways. For example, in addition to (or instead of) increasing drift rate, RPEs might also increase the learning rate on item-context associations – this would have the effect of stamping in the episodic memory of the high RPE event more strongly. More simulation work is needed to determine what combination of mechanisms does the best overall job of explaining the effects of RPEs on declarative memory.

High RPEs are better remembered

Consistent with previous results (Rouhani et al., 2018), we found that high RPEs led to better recognition memory for the event associated with the RPE. This finding is moreover consistent with work showing enhanced memory for other types of surprising events in the context of reward learning (Murty & Adcock, 2014; Murty, Labar, & Adcock, 2016), and outside of reward learning (Greve et al., 2017; Kalbe & Schwabe, 2019; De Loof et al., 2018; Ergo et al., 2019).

High RPEs form event boundaries in memory

We found that latent shifts in the reward value of a rewarding source induce event boundaries by interrupting the sequential integration of memories that occur before and after a high-RPE event, thus acting similarly to other event boundaries reported in the literature (DuBrow & Davachi, 2013, 2014; Ezzyat & Davachi, 2014; Horner et al., 2016). Heusser and colleagues recently demonstrated that enhanced associative memory for a perceptual boundary comes at the cost of integrating events across the boundary, reflecting a trade-off between the two processes (Heusser et al., 2018). Here, we found concordant results in the domain of latent reward expectations: high-RPE events were not only better encoded but also demonstrated intact associative memory with their preceding items, through intact recognition priming and sequence memory. However, and in line with this trade-off, events surrounding the high RPE demonstrated diminished associative memory through impaired recognition priming and sequence memory.

Across all experiments, with the exception of the distance judgment results in Experiment 4 (which need to be replicated), we did not find effects of the sign or a continuous measure of the RPE on our behavioral measures. Previous work shows that positive RPEs increase memory for associated events more than negative RPEs (Jang et al., 2019), and that the strength of the RPE, regardless of sign, enhances memory for events experienced during reward learning (Rouhani et al., 2018). Here, large latent jumps in rewards, regardless of their sign and exact size, modulated the structure of memory, implicating neural processes associated with event boundaries rather than reward learning per se (see 'Neural mechanisms' below). We note that, in our task, we were interested in the effect of having detected a change in rewards (i.e., to induce sharp event boundaries), and so the jumps in the underlying reward distribution were quite obvious. In the real world, however, these changes may be more subtle and gradual, requiring multiple observations to infer an event boundary. Future work could introduce uncertainty around reward shifts and examine how this affects the temporal organization of events in memory (DuBrow et al., 2017).

Another key issue is whether the (apparent) contextual discontinuity evoked by high-RPE events in our study is attributable to the prediction error *per se*, or whether it is attributable to the fact that high RPEs indicated shifts in the underlying 'latent cause' driving participants' observations (Zacks et al., 2007). In our paradigm, these two factors (RPE and shift-in-latent-cause) were confounded – in future work, we can try to unconfound them (e.g, by having isolated high-reward or low-reward items that do not indicate a lasting change in the underlying mean reward value). Related to this point, Siefke et al. recently ran a study that attempted to unconfound context change and prediction error, using stimuli that varied in their background color; results from that study supported the hypothesis that context change, not prediction error *per se*, is the key determinant of discontinuities in mental context (Siefke et al., 2019). More work is needed to see if this applies to our RPE paradigm.

Event boundary occurs across the high-RPE event

Although some theories (e.g., latent cause models; Gershman et al., 2014) predict that an event boundary occurs at the high prediction error event itself, separating that event from preceding items, we found intact associative links between the high-RPE event and its predecessor. At the same time, we found evidence for an event boundary *across* the high-RPE event. In our model, the high-RPE item and its predecessor are linked because the high-RPE$_{-1}$ item is active in the context layer when the high-RPE item is presented at study. Additionally, the high-RPE item is strongly linked to its successor since the high-RPE item itself gets strongly inserted into the high-RPE_{+1} item's context. For this reason, although we did not test for recognition priming between the high-RPE item and high-RPE_{+1} item, we predict, based on our model, that there will be strong recognition priming for the high-RPE_{+1} item when primed by the high-RPE item. Nevertheless, consistent with our behavioral results, the model predicts that recognition priming between the high-RPE_{-1} and the high-RPE_{+1} item will be disrupted because the increased drift associated with the high RPE leads to weak representation of the high-RPE_{-1} item in the high-RPE_{+1}'s context. In sum, our model predicts that the high-RPE item is linked to both its predecessor and successor through context while disrupting the association of the events around it. This explains the seemingly inconsistent results of our Experiments 1 and 2 (and the '0 between' condition in Exp. 4), which suggest no boundary between the high-RPE item and its predecessor, and Experiments 3 and 4 that provide evidence for a boundary across the high-RPE event.

Previous work offers another potential mechanism for the preserved link between the high-RPE event and its preceding event, namely that at event boundaries, memory of the previous episode is reinstated (Sols et al., 2017), perhaps leading to binding between the high-RPE event and its predecessor. Other work has also shown that increased hippocampal activity at event offset (i.e., right after the boundary is inferred) predicts subsequent retrieval of the previous episode, in a sense 'registering' the just-experienced episode (Baldassano et al., 2017; Ben-Yakov & Dudai, 2011; Ben-Yakov et al., 2013). In our task, the boundary itself is calculated by the difference between the expected value and the current reward, which, along with the 'replay' mechanism described above, could additionally bind the high-RPE event with its predecessor.

Recognition priming for high- versus low-RPE events

Our results in Experiments 1 and 2 were suggestive of more stable recognition priming for high-RPE items than for low-RPE items (i.e., numerically, high-RPE items were retrieved faster than the low-RPE targets, although not significantly). This pattern was also present in the model, where high-RPE items were more strongly associated with their predecessor than low-RPE items. We note that in previous studies, recognition priming was evident only for high-confidence recognition (i.e., for recollection instead of familiarity; DuBrow & Davachi, 2013; Schwartz et al., 2005), and we did not collect confidence judgments in our task, perhaps occluding more stable recognition priming effects in the low-RPE pairs. If anything, however, this emphasizes the intact association of the high-RPE event with its predecessor, as we saw recognition priming for the high-RPE item across all confidence levels.

Sequence memory

We found two distinct effects of our RPE conditions on sequence memory: (1) temporal order memory was impaired for items spanning a high (compared to low) RPE event, and (2) temporal order memory in the low-RPE condition (but not the high-RPE condition) was enhanced with greater distance between the events at encoding. In our simulation, primacy judgments were based on which item's context was more distant from the test context (Friedman, 1993, 2004; Hintzman, 2002). Using this mechanism, our simulation captured the second effect (increased accuracy with greater distance between events at encoding, in the low-RPE condition) but it failed to predict the first effect (impaired sequence memory for items spanning a high RPE event). This is because high-RPE events in our model create a contextual shift, thereby making it easier (instead of harder) to identify the first item as the more contextually-distant (and thus earlier) item.

These results suggest that different mechanisms support temporal order memory for items within a latent context versus those across one (DuBrow & Davachi, 2016; Ezzyat & Davachi, 2014). Here, we suggest that a mechanism relying on contextual distance (i.e., the present simulation) can explain temporal memory for items within a latent context, but not temporal memory across latent contexts. As discussed above, a model incorporating an explicit reconstruction mechanism like SEM (Franklin et al., 2020) may do better at explaining the full pattern of results.

Distance memory

Event boundaries increase the subjective temporal distance between events (Ezzyat & Davachi, 2014). We saw this effect only in the replication of Experiment 4, which points to more variable results with this measure. Across both datasets in Experiment 4, participants' responses were quite inaccurate: the mean distance judgment was close to '3 between,' even though two-thirds of the actual distances were smaller than 3. Although this needs to be replicated, we did find greater subjective distance for high-RPE pairs that was largely driven by the pair that included the high-RPE event itself ('0 between'). However, this condition did not demonstrate impaired sequence memory, which points to a dissociation between mechanisms supporting sequence and distance memory. This finding is in line with a recent study showing that, at event boundaries, separate components of pupillary response are associated with sequence and distance memory (Clewett et al., 2019).

Neural mechanisms

RPEs modulate dopamine release in the ventral tegmental area (VTA) by increasing firing when rewards are better than expected, and decreasing firing when rewards are worse than expected (Barto, 1995; Montague et al., 1996). Given dopamine-dependent plasticity in the hippocampus, associated with memory formation, putative links have been made between RPE signals in the VTA and modulation of hippocampal plasticity (Lisman & Grace, 2005), giving rise to enhanced memory for events that are better than expected (Jang et al., 2019).

In this study, however, we only observed effects of unsigned RPEs on the structure of memory. Recent work offers a mechanism by which unsigned (absolute value) RPEs can interact with memory. The locus coeruleus (LC), a previously unknown source of dopamine, co-releases dopamine along with its known release of norepinephrine, facilitating the generation of hippocampal memories during learning and for novel events (Kempadoo et al., 2016; Takeuchi et al., 2016). Large RPEs, whether positive or negative, have been shown to increase learning rate during reward learning, and are thought to modulate the noradrenergic LC system and its connections to the anterior cingulate cortex (Behrens et al., 2007; Courville, Daw, & Touretzky, 2006; Nassar et al., 2012; Roesch et al., 2012; Sara, 2009) – a system linked to memory for surprising or arousing events (Clewett et al., 2018, 2014). Moreover, an increase in pupil dilation (a biomarker for LC activation) occurs at event boundaries, and predicts sequence and distance memory (Clewett et al., 2019), providing further corroboration of the putative role of the LC in supporting our results.

Nevertheless, there is still a question of whether this LC mechanism *strengthens* the high-RPE event in memory and/or *segments* it from previous events. For example, it is possible that LC enhances the encoding of the high-RPE event while the hippocampus segments or 'pattern separates' the events that come after the high RPE from those that came before (Yassa & Stark, 2011). Future work should characterize how these potentially distinct processes of strengthening and segmenting are orchestrated by the brain.

Moreover, prediction errors are thought to enact a ‘network reset’ (Zacks et al., 2007) that has been recently linked to a shifting latent-state representation in the orbitofrontal cortex (Nassar, McGuire, Ritz, & Kable, 2018). The orbitofrontal cortex is a strong candidate region for representing these latent states (Schuck, Cai, Wilson, & Niv, 2016), which are thought to encode a cognitive map of task space (Wilson, Takahashi, Schoenbaum, & Niv, 2014). Seeing that event boundaries modulate representations in the hippocampus (DuBrow & Davachi, 2014; Ezzyat & Davachi, 2014), it has been suggested that at these boundaries, enhanced hippocampal activity and a shift in cortical representations (such as in the orbitofrontal cortex) increases the drift in temporal context (Brunec, Moscovitch, & Barense, 2018). Future work should characterize the interactions between the orbitofrontal cortex and the hippocampus in segmenting our experiences and organizing those memories.

SUMMARY

Using four experiments, we established that latent shifts in the mean value of a reward distribution, generating the experience of high reward prediction errors, led to stronger recognition for the event associated with the high prediction error and preserved its link to the preceding event, while simultaneously interrupting the sequential integration of events across the prediction error event, thereby creating an event boundary in memory. We developed a computational model that treats a high prediction error event as an increase in the updating of that event to an internal, temporal context during encoding (thus creating a representational break between the events that occurred before and after the high prediction error event), and were able to capture our recognition priming and sequence memory results. These results suggest that large changes in the value of a rewarding experience split our memories of those experiences, separating them into separate clusters in memory, each including similarly rewarding events. This mechanism can help create low-dimensional representations of task states that are useful for both learning and decision making.

5

Conclusion

Reinforcement learning and episodic memory have been traditionally studied separately, in part because of their association with distinct neural mechanisms (Shohamy & Adcock, 2010), yet both systems are critical to adaptive decision-making. In fact, more recent and converging work from psychology, neuroscience, and computer science has considerably advanced our understanding of how the brain accomplishes learning by characterizing the interaction between these two systems (for a review, see Gershman & Daw, 2017). In this dissertation, I investigated the role of reward prediction errors (RPEs), the reinforcement-learning signal that drives trial-and-error learning, in shaping our episodic memory for those experiences. Below, I bridge across the results of our studies by offering putative neural mechanisms and addressing some remaining puzzles.

NEURAL MECHANISMS

Across our studies, we found that large unsigned RPEs ('large RPEs') experienced at reward outcome increased both the learning of values (learning rate) and memory for those events. Large RPEs enhance attention (Pearce & Hall, 1980), supported by increased firing of noradrenergic neurons from the locus coeruleus (LC; Sara, 2009) which further project to the amygdala (Holland & Schiffino, 2016; Roesch, Calu, Esber, & Schoenbaum, 2010) and the anterior cingulate cortex (Roesch et al., 2012) – all regions linked to the unsigned-RPE modulation of learning rate (Behrens et al., 2007; McGuire et al., 2014; Nassar et al., 2012; Pearce & Hall, 1980). Although the effects of unsigned RPEs on memory have been less studied, prediction errors, more generally, are an established signal in modulating memory. They have been associated with the effects of surprise (e.g., Greve et al., 2017), novelty or more broadly, any change event (Zacks et al., 2007), on memory. Distinguishing the mechanisms supporting these different kinds of prediction errors is an important direction for future research (Schomaker & Meeter, 2015; Duncan & Schlichting, 2018). Regardless, they may each share the deployment of the LC-norepinephrine (LC-NE) system at the prediction-error event, putatively giving rise to both noradenergic and dopaminergic modulation of hippocampal memories (Kempadoo et al., 2016; Takeuchi et al., 2016; Wagatsuma et al., 2017).

We also showed evidence of a signed-RPE effect on memory (Chapter 3), consistent with the canonical dopamine signal from the midbrain, which increases when rewards are better than expected and decreases when they are worse than expected (Barto, 1995; Montague et al., 1996). We found this signed, memory-enhancing RPE to occur for reward-predicting cues instead of events at reward outcome. This effect is putatively supported by the dopamine signal traveling from reward outcome to the cue predicting outcome with more learning (Schultz et al., 1997). This finding is consistent with work showing better memory for events that elicit high-reward expectation and anticipation (Stanek et al., 2019; Murty & Adcock, 2014; Adcock et al., 2006; Jang et al., 2019; Wittmann et al., 2005), where increased dopaminergic firing at cue amplifies the hippocampal encoding of events (Murty & Adcock, 2014). More recently, enhanced hippocampal activation during reward anticipation has been speculated to reflect the simulation of future rewards, thus proposing that the hippocampus is engaging in a retrieval process (i.e., the simulation or 'retrieval' of future outcomes) versus an encoding process. Given that the hippocampus shifts between retrieval and encoding states (Hasselmo et al., 2002; Duncan et al., 2012), future work should identify hippocampal dynamics during reward anticipation.

Regardless, the role of hippocampal retrieval during reinforcement learning is of growing interest, as demonstrated by episodic sampling models of decision-making (Lieder et al., 2018; Shadlen & Shohamy, 2016; Bakkour et al., 2019; Bornstein & Norman, 2017; Bornstein et al., 2017). Unlike standard reinforcement learning where values are generated by a recency-weighted average over all experienced rewards, these episodic values rely on individual memories of past outcomes, or a distribution of those memories, that are sampled online when deliberating between options. These models thus allow for the prioritization of particular memories over others during decision-making, thereby biasing choice. The better memory for large-RPE events observed in our experiments suggests that these events may be prioritized during decision-making. Accordingly, a new 'utility-weighted sampling' model increases the sampling of extreme (i.e., large-RPE) outcomes to predict a wide-range of choice behavior, and further demonstrates that an over-representation of extreme outcomes in memory is optimal when making decisions under constraints (Lieder et al., 2018).

Greater sampling of large-RPE events can also occur offline through replay and consolidation of those events in memory. Although all of the effects reported here were tested shortly after encoding, large RPEs elicit the arousal and dopaminergic firing thought to predict increased replay and consolidation of those events (Mattar & Daw, 2018; Momennejad, Otto, Daw, & Norman, 2018; Russek, Momennejad, Botvinick, Gershman, & Daw, 2017). I therefore hypothesize that our effects would be observable and potentially strengthened after consolidation, in line with work showing signed-RPE effects on memory after both a short and 24-hour delay (Jang et al., 2019).

Distinct Unsigned-RPE Effects on Learning and Memory

In Chapter 1, we first found large RPEs to increase both learning rate and memory, but we did not find learning rate and memory to be correlated. Similarly, in the experiments of Chapter 3, large (outcome) RPEs again increased learning rate and memory, but learning rates did not predict memory for the outcome events that had generated that value update, nor did they predict memory for the subsequent cue event where the value update was applied. These results held true for large RPEs caused by either high-outcome variance or by change-point events. This consistent lack of relationship between learning rate for values and memory for those events, raises the question of how large RPEs may be differentially influencing learning and memory.

Chapter 1 results provide a strong illustration of this dissociation. In these experiments, participants learned to predict rewards in contexts characterized by 'high' or 'low risk' (i.e., high or low-outcome variance). We found that learning rate scaled with the reward variance of the context whereas memory did not. In line with previous work (Diederen et al., 2016; Diederen & Schultz, 2015; Nassar, Wilson, Heasly, & Gold, 2010), learning rates were, in general, lower in the high-risk context, and therefore less sensitive to the larger fluctuations of reward outcomes, demonstrating adaptive scaling to the expected uncertainty of that context. Memory, however, was higher in the high-risk context thus showing no such adaptation. These results may suggest that although unsigned RPEs drive both increases in learning rate and memory, an additional step occurs for learning rates, where estimated values reflect a scaling of prediction errors by the uncertainty of the reward context, which may in effect decouple its relationship to memory.

Such scaling has been observed for the reward prediction error itself, where dopaminergic firing adapts to the uncertainty of the reward environment by showing greater sensitivity to outcomes in a low- versus high-variance reward context (Tobler et al., 2005). Another possibility, is that learning-rate adaptation is influenced by acetylcholine, whose role in signalling the 'expected uncertainty' of a learning environment (i.e., its reward variance) has been contrasted with the noradrenergic signalling of 'unexpected uncertainty' (changes in the reward state, i.e., change-point events; Yu & Dayan, 2005). A higher level of acetylcholine is associated with greater expected uncertainty (as in a high-risk context). Interestingly, increased acetylcholine is also thought to orchestrate hippocampal pattern separation, giving rise to enhanced encoding of experienced events (Hasselmo et al., 2002; Duncan & Schlichting, 2018). Nevertheless, very little is known about how dopamine, norepinephrine, and acetylcholine may independently or interactively contribute to value updating and memory for those events, begging further investigation.

Moreover, this dissociation between learning rate and memory could be potentially explained by an alternative model of decision-making, episodic sampling, where these seemingly contradictory effects may in fact be complementary. In our paradigm, it's possible that a sample of individual outcomes was being recruited from memory when estimating the value of that reward cue. Accordingly, in the high-risk context, where memory was better for large-RPE events, a greater number of more active outcome memories could have led participants to average across more outcomes from the past when updating their values, which is equivalent to a lower learning rate.* Higher learning rates in the low-risk context, on the other hand, reflect the integration of fewer events from the past, consistent with worse memory for those less surprising events.

Although we cannot adjudicate between a scaling versus sampling account for the observed dissociation between learning rate and memory, a neural investigation probing the scaling of prediction errors during reward outcome (e.g., Diederen et al., 2016) versus memory sampling at reward cue (e.g., Bakkour et al., 2019) may shed light on the underlying mechanism supporting these separate effects.

The Strengthening and Segmenting of Memories

In our experiments, we investigated the mnemonic effects of large RPEs generated by either high outcome variance (e.g., high-risk context) or by a change in the mean of the underlying reward distribution (change-point event). Regardless of this difference, large RPEs boosted recognition memory for those events across all of our studies. However, when testing memory for the temporal structure of events, this difference led to an opposite pattern of results. Importantly, a change-point event indicates a shift in the latent reward context (DuBrow et al., 2017), thereby creating an event boundary in memory that segments experiences occurring before the change point from those that came after it. In Chapter 4 , we showed evidence of an RPE-induced event boundary through impaired recognition priming and sequence memory for events across a change-point (Rouhani et al., 2020). On the other hand, when we tested whether large RPEs caused by high-outcome variance affected the structure of memory (Chapter 1; Rouhani et al., 2018), we in fact found better sequence memory for items within a high versus low-risk context. These two seemingly contradictory results indicate that large RPEs alone do not create event boundaries in memory, they need to further reflect a (meaningful) change in the la-

---

*Lower learning rates indicate that a smaller proportion of the most recent outcome is used to update the internal value re of a reward cue, thereby relying more on past outcomes.

tent reward context. Otherwise, a reward context characterized by high-outcome variance – where large RPEs potentially increase attention and memory for more events – gives rise to better memory for the temporal organization of events.

It is furthermore unclear how the LC-NE system may help in the strengthening and/or segmenting of these events in memory. These two effects may both rely on the LC-NE system, where strengthening and segmenting occur at different time-points. Evidence for this comes from a study showing that a surprising event predicted an early increase in pupil dilation, a marker of norephineprine release, whereas an event indicating a model update (i.e., a change-point event) predicted a late decrease in pupil dilation (O'Reilly et al., 2013). In our paradigm, it is possible that a change-point event first increased noradrenergic firing thereby strengthening that event in memory, and then decreased it thus engaging in memory segmentation. In line with this, in our experiments (Chapter 4; Rouhani et al., 2020), we did not find memory segmentation to occur at the change-point event itself but instead found it for events across the change-point. In fact, in our CMR model, memory segmentation needed to occur after the change-point event in order to explain our results (in our model, every event was integrated into the next event's context).

As another possibility, strengthening and segmenting could be orchestrated by distinct mechanisms. Evidence for this comes from work demonstrating that event segmentation across change points is associated with shifting state representations in the orbitofrontal cortex (Nassar et al., 2018) as well as in the hippocampus (DuBrow & Davachi, 2014; Ezzyat & Davachi, 2014). We would not expect large RPEs experienced in the same latent context (i.e., in a high-risk context) to induce such representational changes. A neural investigation of these distinct effects of large RPEs on the structure of memory could potentially disentangle the mechanisms supporting the mnemonic strengthening and segmenting elicited by large-RPE events.

To conclude, I show that the signal driving reinforcement learning, the reward prediction error, dynamically influences learning while separately strengthening and, at times, segmenting our memory of those events, thereby offering neuroscientific targets for future theoretical and empirical work.

References

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. (2006). Reward-Motivated Learning: Mesolimbic Activation Precedes Memory Formation. *Neuron*, *50*(3), 507–517. doi: 10.1016/j.neuron.2006.03.036

Admon, R., Nickerson, L. D., Dillon, D. G., Holmes, A. J., Bogdan, R., Kumar, P., ... Pizzagalli, D. A. (2015). Dissociable cortico-striatal connectivity abnormalities in major depression in response to monetary gains and penalties. *Psychological Medicine*, *45*(1), 121–131. doi: 10.1017/S0033291714001123

Admon, R., & Pizzagalli, D. A. (2015). Dysfunctional reward processing in depression. *Current Opinion in Psychology*, *4*, 114–118. doi: 10.1016/j.copsyc.2014.12.011

Bakkour, A., Palombo, D. J., Zylberberg, A., Kang, Y. H., Reid, A., Verfaellie, M., ... Shohamy, D. (2019). The hippocampus supports deliberation during value-based decisions. *eLife*, *8*, 1–28. doi: 10.7554/eLife.46080

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017, 8). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, *95*(3), 709–721. doi: 10.1016/j.neuron.2017.06.041

Barto, A. G. (1995). Adaptive critic and the basal ganglia. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (p. 215). MIT press.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... Grothendieck, G. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Beesley, T., Nguyen, K. P., Pearson, D., & Le Pelley, M. E. (2015). Uncertainty and predictiveness determine attention to cues during human associative learning. *Quarterly Journal of Experimental Psychology*, *68*(11), 2175–2199. doi: 10.1080/17470218.2015.1009919

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–21. doi: 10.1038/nn1954

Ben-Yakov, A., & Dudai, Y. (2011). Constructing Realistic Engrams: Poststimulus Activity of Hippocampus and Dorsal Striatum Predicts Subsequent Episodic Memory. *Journal of Neuroscience*, *31*(24), 9032–9042. doi: 10.1523/jneurosci.0702-11.2011

Ben-Yakov, A., Eshel, N., & Dudai, Y. (2013). Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. *Journal of Experimental Psychology: General*, *142*(4), 1255–1263. doi: 10.1037/a0033558

Biele, G., Erev, I., & Ert, E. (2009). Learning, risk attitude and hot stoves in restless bandit problems. *Journal of Mathematical Psychology*, *53*, 155–167. doi: 10.1016/j.jmp.2008.05.006

Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017, 6). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, *8*, 15958. doi: 10.1038/ncomms15958

Bornstein, A. M., & Norman, K. A. (2017, 6). Reinstated episodic context guides sampling-based decisions for reward. *Nature Neuroscience*, *20*(7), 997–1003. doi: 10.1038/nn.4573

Brunec, I. K., Moscovitch, M., & Barense, M. D. (2018, 7). Boundaries Shape Cognitive Representations of Spaces and Events. *Trends in Cognitive Sciences*, *22*(7), 637–650. doi: 10.1016/j.tics.2018.03.013

Burt, D. B., Zembar, M. J., & Niederehe, G. (1995). Depression and memory impairment: A meta-analysis of the association, its pattern, and specificity. *Psychological Bulletin*, *117*(2), 285–305. doi: 10.1037/0033-2909.117.2.285

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1491 - 1494. doi: 10.1037/0278-7393.27.6.1491

Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning*, *1*(1), 69–84. doi: 10.1007/s11409-006-6894-z

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017, 1). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. doi: 10.18637/jss.v076.i01

Chase, H. W., Frank, M. J., Michael, A., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2010). Approach and avoidance learning in patients with major depression and healthy controls: Relation to anhedonia. *Psychological Medicine*, *40*(3), 433–440. doi: 10.1017/S0033291709990468

Clewett, D., Gasser, C., & Davachi, L. (2019). Dynamic arousal signals construct memories of time and events. *bioRxiv*. doi: 10.1101/765214

Clewett, D., Huang, R., Velasco, R., Lee, T.-H., & Mather, M. (2018). Locus coeruleus activity strengthens prioritized memories under arousal. *The Journal of Neuroscience*, 2097–17. doi: 10.1523/JNEUROSCI.2097-17.2017

Clewett, D., & Murty, V. P. (2019). Echoes of emotions past: How neuromodulators determine what we recollect. *eNeuro*, *6*(2). doi: 10.1523/ENEURO.0108-18.2019

Clewett, D., Schoeke, A., & Mather, M. (2014). Locus coeruleus neuromodulation of memories encoded during negative or unexpected action outcomes. *Neurobiology of Learning and Memory*, *111*, 65–70. doi: 10.1016/j.nlm.2014.03.006

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*(7), 294–300. doi: 10.1016/j.tics.2006.05.004

Davachi, L., & Dubrow, S. (2015). How the hippocampus preserves order: the role of prediction and context. *Trends in Cognitive Sciences*, *19*(2), 92–99. doi: 10.1016/j.tics.2014.12.004

Davidow, J. Y., Foerde, K., Galvan, A., & Shohamy, D. (2016). An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence. *Neuron*, *92*(1), 93–99. doi: 10.1016/j.neuron.2016.08.031

De Loof, E., Ergo, K., Naert, L., Janssens, C., Talsma, D., Van Opstal, F., & Verguts, T. (2018). Signed reward prediction errors drive declarative learning. *PLoS ONE*, *13*(1). doi: 10.1371/journal.pone.0189212

Diederen, K. M. J., & Schultz, W. (2015). Scaling prediction errors to reward variability benefits error-driven learning in humans. *Journal of Neurophysiology*, *114*, 1628–1640.

Diederen, K. M. J., Spencer, T., Vestergaard, M. D., Fletcher, P. C., & Schultz, W. (2016). Adaptive Prediction Error Coding in the Human Midbrain and Striatum Facilitates Behavioral Adaptation and Learning Efficiency. *Neuron*, *90*, 1127–1138. doi: 10.1016/j.neuron.2016.04.019

Difallah, D. E., Catasta, M., Demartini, G., Ipeirotis, P. G., & Cudré-Mauroux, P. (2015). The Dynamics of Micro-Task Crowdsourcing: The Case of Amazon MTurk. In *Proceedings of the 24th international conference on world wide web* (pp. 238–247). Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee. doi: 10.1145/2736277.2741685

Dillon, D. G., Dobbins, I. G., & Pizzagalli, D. A. (2014). Weak Reward Source Memory in Depression Reflects Blunted Activation of VTA/SN and Parahippocampus. *Social Cognitive and Affective Neuroscience*, *9*(10), 1576–1583.

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, *142*(4), 1277–86. doi: 10.1037/a0034024

DuBrow, S., & Davachi, L. (2014). Temporal Memory Is Shaped by Encoding Stability and Intervening Item Reactivation. *Journal of Neuroscience*, *34*(42), 13998–14005. doi: 10.1523/jneurosci.2535-14.2014

DuBrow, S., & Davachi, L. (2016). Temporal binding within and across events. *Neurobiology of Learning and Memory*, *134*, 107–114. doi: 10.1016/j.nlm.2016.07.011

DuBrow, S., Rouhani, N., Niv, Y., & Norman, K. A. (2017). Does mental context drift or shift? *Current Opinion in Behavioral Sciences*, *17*, 141–146. doi: 10.1016/j.cobeha.2017.08.003

Duncan, K. D., Sadanand, A., & Davachi, L. (2012). Memory's Penumbra: Episodic Memory Decisions Induce Lingering Mnemonic Biases. *Science*, *337*.

Duncan, K. D., & Schlichting, M. L. (2018). Hippocampal representations as a function of time, subregion, and brain state. *Neurobiology of Learning and Memory*, *153*, 40–56. doi: 10.1016/j.nlm.2018.03.006

Duncan, K. D., & Shohamy, D. (2016). Memory States Influence Value-Based Decisions. *Journal of Experimental Psychology: General*, *145*(9), 3–9.

Duszkiewicz, A. J., Mcnamara, C. G., Takeuchi, T., & Genzel, L. (2018). Novelty and Dopaminergic Modulation of Memory Persistence: A Tale of Two Systems. *Trends in Neurosciences*. doi: 10.1016/j.tins.2018.10.002

Eldar, E., & Niv, Y. (2015). Interaction between emotional state and learning underlies mood instability. *Nature Communications*, *6*. doi: 10.1038/ncomms7149

Eldar, E., Rutledge, R. B., Dolan, R. J., & Niv, Y. (2016). *Mood as Representation of Momentum* (Vol. 20) (No. 1). doi: 10.1016/j.tics.2015.07.010

Ergo, K., De Loof, E., Janssens, C., & Verguts, T. (2019). Oscillatory signatures of reward prediction errors in declarative learning. *NeuroImage*, *186*, 137–145. doi: 10.1016/j.neuroimage.2018.10.083

Ergo, K., De Loof, E., & Verguts, T. (2020). Reward Prediction Error and Declarative Memory. *Trends in Cognitive Sciences*, *24*(5), 388–397. doi: 10.1016/j.tics.2020.02.009

Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, *81*(5), 1179–1189. doi: 10.1016/j.neuron.2014.01.042

Foerde, K., Braun, E. K., & Shohamy, D. (2012). A Trade-Off between Feedback-Based Learning and Episodic Memory for Feedback Events: Evidence from Parkinson's Disease. *Neurodegenerative Diseases*, *11*(2), 93–101. doi: 10.1159/000342000

Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience*, *5*(5), 458–462. doi: 10.1038/nn834

Franklin, N. T., Norman, K. A., Ranganath, C., Zacks, J. M., & Gershman, S. J. (2020). Structured Event Memory: A neuro-symbolic model of event cognition. *Psychological Review*, *127*(3), 327.

Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin*, *113*(1), 44–66. doi: 10.1037/0033-2909.113.1.44

Friedman, W. J. (2004). Time in autobiographical memory. *Social Cognition*, *22*(5), 591–605. doi: 10.1521/soco.22.5.591.50766

Gaddy, M. A., & Ingram, R. E. (2014). A meta-analytic review of mood-congruent implicit memory in depressed mood. *Clinical Psychology Review*, *34*(5), 402–416. doi: 10.1016/j.cpr.2014.06.001

Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology*, *68*, 101–128. doi: 10.1146/annurev-psych-122414-033625

Gershman, S. J., & Niv, Y. (2013). Perceptual estimation obeys Occam's razor. *Frontiers in Psychology*, *4*, 623. doi: 10.3389/fpsyg.2013.00623

Gershman, S. J., Radulescu, A., Norman, K. A., & Niv, Y. (2014). Statistical Computations Underlying the Dynamics of Memory Updating. *PLoS Computational Biology*, *10*(11). doi: 10.1371/journal.pcbi.1003939

Gotlib, I. H., & Joormann, J. (2010). Cognition and Depression: Current Status and Future Directions. *Annual Review of Clinical Psychology*, *6*(1), 285–312. doi: 10.1146/annurev.clinpsy.121208.131305

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., ... Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain*, *134*(6), 1751–1764. doi: 10.1093/brain/awr059

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, *94*, 149–165. doi: 10.1016/j.jml.2016.11.001

Hasselmo, M. E., Bodelón, C., & Wyble, B. P. (2002, 4). A Proposed Function for Hippocampal Theta Rhythm: Separate Phases of Encoding and Retrieval Enhance Reversal of Prior Learning. *Neural Computation*, *14*(4), 793–817. doi: 10.1162/089976602317318965

Hauser, T. U., Eldar, E., Purg, N., Moutoussis, M., & Dolan, R. J. (2019). Distinct Roles of Dopamine and Noradrenaline in Incidental Memory. *The Journal of Neuroscience*, *39*(39), 7715–7721. doi: 10.1523/JNEUROSCI.0401-19.2019

Heusser, A. C., Ezzyat, Y., Shiff, I., & Davachi, L. (2018). Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*(212), 1–48. doi: http://dx.doi.org/10.1037/xlm0000503

Hintzman, D. L. (2002). Context matching and judgments of recency. *Psychonomic Bulletin and Review*, *9*(2), 368–374. doi: 10.3758/BF03196295

Holland, P. C., & Schiffino, F. L. (2016). Mini-review: Prediction errors, attention and associative learning. *Neurobiology of Learning and Memory*, *131*, 207–215. doi: 10.1016/j.nlm.2016.02.014

Horner, A. J., Bisby, J. A., Wang, A., Bogus, K., & Burgess, N. (2016). The role of spatial boundaries in shaping long-term event representations. *Cognition*, *154*, 151–164. doi: 10.1016/j.cognition.2016.05.013

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*(3), 269–299. doi: 10.1006/jmps.2001.1388

Hsu, M., Krajbich, I., Zhao, C., & Camerer, C. F. (2009). Neural Response to Reward Anticipation under Risk Is Nonlinear in Probabilities. *Journal of Neuroscience*, *29*(7), 2231–2237. doi: 10.1523/JNEUROSCI.5296-08.2009

Hulley, S. B. (2007). *Designing clinical research*. Lippincott Williams & Wilkins.

Huys, Q. J., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biology of Mood & Anxiety Disorders*, *3*(1), 12. doi: 10.1186/2045-5380-3-12

Iigaya, K., Hauser, T., Kurth-Nelson, Z., O'Doherty, J., Dayan, P., & Dolan, R. (2019). The value of what's to come: neural mechanisms coupling prediction error and reward anticipation. *bioRxiv*, 588699. doi: 10.1101/588699

Ipeirotis, P. G. (2010, 12). Analyzing the Amazon Mechanical Turk Marketplace. *XRDS*, *17*(2), 16–21. doi: 10.1145/1869086.1869094

Jang, A. I., Nassar, M. R., Dillon, D. G., & Frank, M. J. (2019). Positive reward prediction errors during decision-making strengthen memory encoding. *Nature Human Behaviour*, *3*(7), 719–732. doi: 10.1038/s41562-019-0597-3

Joormann, J., & Quinn, M. E. (2014). Cognitive processes and emotion regulation in depression. *Depression and Anxiety*, *31*(4), 308–315. doi: 10.1002/da.22264

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291. doi: 10.1017/CBO9781107415324.004

Kalbe, F., & Schwabe, L. (2019). Beyond Arousal: Prediction Error Related to Aversive Events Promotes Episodic Memory Formation. *Journal of Experimental Psychology: Learning Memory and Cognition*. doi: 10.1037/xlm0000728

Kempadoo, K. A., Mosharov, E. V., Choi, S. J., Sulzer, D., & Kandel, E. R. (2016). Dopamine release from the locus coeruleus to the dorsal hippocampus promotes spatial learning and memory. *Proceedings of the National Academy of Sciences*, *113*(51). doi: 10.1073/pnas.1616515114

Kesner, R. P., Hunsaker, M. R., & Ziegler, W. (2010, 1). The role of the dorsal CA1 and ventral CA1 in memory for the temporal order of a sequence of odors. *Neurobiology of Learning and Memory*, *93*(1), 111–116. doi: 10.1016/j.nlm.2009.08.010

Knutson, B., Bhanji, J. P., Cooney, R. E., Atlas, L. Y., & Gotlib, I. H. (2008). Neural Responses to Monetary Incentives in Major Depression. *Biological Psychiatry*, *63*(7), 686–692. doi: 10.1016/j.biopsych.2007.07.023

Kragel, J. E., Morton, N. W., & Polyn, S. M. (2015). Neural activity in the medial temporal lobe reveals the fidelity of mental time travel. *Journal of Neuroscience*, *35*(7), 2914–2926. doi: 10.1523/JNEUROSCI.3378-14.2015

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Kumar, P., Goer, F., Murray, L., Dillon, D. G., Beltzer, M. L., Cohen, A. L., ... Pizzagalli, D. A. (2018). Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. *Neuropsychopharmacology*, *43*(7), 1581–1588. doi: 10.1038/s41386-018-0032-x

Leal, S. L., Tighe, S. K., Jones, C. K., & Yassa, M. A. (2014). Pattern separation of emotional information in hippocampal dentate and CA3. *Hippocampus*, *24*(9), 1146–1155. doi: 10.1002/hipo.22298

Le Pelley, M. E. (2004). *The role of associative history in models of associative learning: A selective review and a hybrid model* (Vol. 57) (No. 3). doi: 10.1080/02724990344000141

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological Review*, *125*(1), 1–32. doi: 10.1037/rev0000074

Lisman, J. E., & Grace, A. A. (2005). The hippocampal-VTA loop: Controlling the entry of information into long-term memory. *Neuron*, *46*(5), 703–713. doi: 10.1016/j.neuron.2005.05.002

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review*, *122*(2), 337–363. doi: 10.1037/a0039036

Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2014). Extreme Outcomes Sway Risky Decisions from Experience. *Journal of Behavioral Decision Making*, *27*, 146–156. doi: 10.1002/bdm.1792

Mackintosh, N. J. (1975). A Theory of Attention: Variations in the Associability of Stimuli with Reinforcement. *Psychological Review*, *82*(4), 276–298.

Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic Bulletin & Review*, *21*(3), 629–636. doi: 10.3758/s13423-013-0542-9

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software*, *4*(40), 1541. doi: 10.21105/joss.01541

Marvin, C. B., & Shohamy, D. (2016). Curiosity and reward: Valence predicts choice and information prediction errors enhance learning. *Journal of Experimental Psychology: General*, *145*(3), 266–272. doi: 10.1037/xge0000140

Mather, M., Clewett, D., Sakaki, M., & Harley, C. W. (2015). Norepinephrine ignites local hot spots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *Behavioral and Brain Sciences*, *1*.

Matsumoto, M., & Hikosaka, O. (2007, 6). Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*, *447*(7148), 1111–1115. doi: 10.1038/nature05860

Matt, G. E., Vázquez, C., & Campbell, W. K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review*, *12*(2), 227–255. doi: 10.1016/0272-7358(92)90116-P

Mattar, M. G., & Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, *21*(11), 1609–1617. doi: 10.1038/s41593-018-0232-z

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally Dissociable Influences on Learning Rate in a Dynamic Environment. *Neuron*, *84*(4), 870–881. doi: 10.1016/j.neuron.2014.10.013

Miendlarzewska, E. A., Bavelier, D., & Schwartz, S. (2016). Influence of reward motivation on human declarative memory. *Neuroscience and Biobehavioral Reviews*, *61*, 156–176. doi: 10.1016/j.neubiorev.2015.11.015

Momennejad, I., Otto, A. R., Daw, N. D., & Norman, K. A. (2018). Offline replay supports planning in human reinforcement learning. *eLife*, *7*. doi: 10.7554/eLife.32548

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *The Journal of Neuroscience*, *16*(5), 1936–47.

Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic Memory and Beyond: The Hippocampus and Neocortex in Transformation. *Annual Review of Psychology*, *67*(1), 105–134. doi: 10.1146/annurev-psych-113011-143733

Mueller, E. M., Pechtel, P., Cohen, A. L., Douglas, S. R., & Pizzagalli, D. A. (2015). Potentiated processing of negative feedback in depression is attenuated by anhedonia. *Depression and Anxiety*, *32*(4), 296–305. doi: 10.1002/da.22338

Murty, V. P., & Adcock, R. A. (2014). Enriched encoding: Reward motivation organizes cortical networks for hippocampal detection of unexpected events. *Cerebral Cortex*, *24*(8), 2160–2168. doi: 10.1093/cercor/bht063

Murty, V. P., Labar, K. S., & Adcock, R. A. (2016). Distinct medial temporal networks encode surprise during motivation by reward versus punishment. *Neurobiology of Learning and Memory*, *134*, 55–64. doi: 10.1016/j.nlm.2016.01.018

Nassar, M. R., McGuire, J. T., Ritz, H., & Kable, J. (2018). Dissociable forms of uncertainty-driven representational change across the human brain. *The Journal of Neuroscience*, 1713–18. doi: 10.1523/jneurosci.1713-18.2018

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. doi: 10.1038/nn.3130

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *The Journal of Neuroscience*, *30*(37), 12366–78. doi: 10.1523/JNEUROSCI.0822-10.2010

Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. *Journal of Neuroscience*, *32*(2), 551–562. doi: 10.1523/JNEUROSCI.5498-10.2012

Niv, Y., & Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends in Cognitive Sciences*, *12*(7), 265–272. doi: 10.1016/j.tics.2008.03.006

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2), 329–337. doi: 10.1016/S0896-6273(03)00169-7

O'Reilly, J. X., Schüffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, *110*(38). doi: 10.1073/pnas.1305373110

Pearce, J. M., & Hall, G. (1980). A Model for Pavlovian Learning: Variations in the Effectiveness of Conditioned But Not of Unconditioned Stimuli. *Psychological Review*, *87*(6), 532–552.

Pearce, J. M., & Mackintosh, N. J. (2010). Two theories of attention: A review and a possible integration. In *Attention and associative learning: From brain to behaviour* (pp. 11–39).

Pickles, A. (1985). *An introduction to likelihood analysis* (No. 42). Geo Books.

Pizzagalli, D. A. (2014). Depression, Stress, and Anhedonia: Toward a Synthesis and Integrated Model. *Annual Review of Clinical Psychology*, *10*(1), 393–423. doi: 10.1146/annurev-clinpsy-050212-185606

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science*, *1*(1-12), 125–126. doi: 10.3758/BF03342823

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A Context Maintenance and Retrieval Model of Organizational Processes in Free Recall. *Psychological Review*, *116*(1), 129–156. doi: 10.1037/a0014420

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditoining: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (Eds.), *Classical conditioning ii: Current research and theory* (pp. 64–99). New York, NY, USA: Appleton-Century-Crofts.

Roesch, M. R., Calu, D. J., Esber, G. R., & Schoenbaum, G. (2010). All that glitters ... dissociating attention and outcome expectancy from prediction errors signals. *Journal of Neurophysiology*, *104*(2), 587–595. doi: 10.1152/jn.00173.2010

Roesch, M. R., Esber, G. R., Li, J., Daw, N. D., & Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. *European Journal of Neuroscience*, *35*(7), 1190–1200. doi: 10.1111/j.1460-9568.2011.07986.x

Rouhani, N. (2018). *Reward prediction error effects on sequence memory* (Tech. Rep.).

Rouhani, N., & Niv, Y. (2019). Depressive symptoms bias the prediction-error enhancement of memory towards negative events in reinforcement learning. *Psychopharmacology*, *236*(8), 2425–2435. doi: 10.1007/s00213-019-05322-z

Rouhani, N., Norman, K. A., & Niv, Y. (2018). Dissociable Effects of Surprising Rewards on Learning and Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/xlm0000518

Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2020, 8). Reward prediction errors create event boundaries in memory. *Cognition*. doi: 10.1101/725440

Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychological Medicine*, *26*(03), 477. doi: 10.1017/S0033291700035558

Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Computational Biology*, *13*(9). doi: 10.1371/journal.pcbi.1005768

Rutledge, R. B., Moutoussis, M., Smittenaar, P., Zeidman, P., Taylor, T., Hrynkiewicz, L., ... Dolan, R. J. (2017). Association of Neural and Emotional Impacts of Reward Prediction Errors With Major Depression. *JAMA Psychiatry*, 10–12. doi: 10.1001/jamapsychiatry.2017.1713

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, *111*(33). doi: 10.1073/pnas.1407535111

Sacchet, M. D., Levy, B. J., Hamilton, J. P., Maksimovskiy, A., Hertel, P. T., Joormann, J., ... Gotlib, I. H. (2017). Cognitive and neural consequences of memory suppression in major depressive disorder. *Cognitive, Affective and Behavioral Neuroscience*, *17*(1), 77–93. doi: 10.3758/s13415-016-0464-x

Sahakyan, L., & Smith, J. R. (2014). "A long time ago, in a context far, far away": Retrospective time estimates and internal context change. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(1), 86–93. doi: 10.1037/a0034250

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, *10*. doi: 10.1038/nrn2573

Schomaker, J., & Meeter, M. (2015). Short- and long-lasting consequences of novelty, deviance and surprise on brain and cognition. *Neuroscience and Biobehavioral Reviews*, *55*, 268–279. doi: 10.1016/j.neubiorev.2015.05.002

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, *91*(6), 1402–1412. doi: 10.1016/j.neuron.2016.08.019

Schultz, W. (2015, 7). Neuronal reward and decision signals: From theories to data. *Physiological Reviews*, *95*(3), 853–951. doi: 10.1152/physrev.00023.2014

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*(5306), 1593–1599. doi: 10.1126/science.275.5306.1593

Schwartz, G., Howard, M. W., Jing, B., & Kahana, M. J. (2005). Shadows of the past: Temporal retrieval effects in recognition memory. *Psychological Science*, *16*(11), 898–904. doi: 10.1111/j.1467-9280.2005.01634.x

Schwarz, G. E. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A Context-Based Theory of Recency and Contiguity in Free Recall. *Psychological Review*, *115*(4), 893–912. doi: 10.1037/a0013396

Shadlen, M. N., & Shohamy, D. (2016). Decision Making and Sequential Sampling from Memory. *Neuron*, *90*(5), 927–939. doi: 10.1016/J.NEURON.2016.04.036

Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in Cognitive Sciences*, *14*(10), 464–472. doi: 10.1016/j.tics.2010.08.002

Shohamy, D., & Daw, N. D. (2015). Integrating memories to guide decisions. *Current Opinion in Behavioral Sciences*, *5*, 85–90. doi: 10.1016/j.cobeha.2015.08.010

Siefke, B. M., Smith, T. A., & Sederberg, P. B. (2019). A context-change account of temporal distinctiveness. *Memory and Cognition*, *47*(6), 1158–1172. doi: 10.3758/s13421-019-00925-5

Smoski, M. J., Felder, J., Bizzell, J., Green, S. R., Ernst, M., Lynch, T. R., & Dichter, G. S. (2009). fMRI of alterations in reward selection, anticipation, and feedback in major depressive disorder. *Journal of Affective Disorders*, *118*(1-3), 69–78. doi: 10.1016/j.jad.2009.01.034

Sols, I., DuBrow, S., Davachi, L., & Fuentemilla, L. (2017, 11). Event Boundaries Trigger Rapid Memory Reinstatement of the Prior Events to Promote Their Representation in Long-Term Memory. *Current Biology*, *27*(22), 3499–3504. doi: 10.1016/j.cub.2017.09.057

Stanek, J. K., Dickerson, K. C., Chiew, K. S., Clement, N. J., & Adcock, R. A. (2019). Expected reward value and reward uncertainty have temporally dissociable effects on memory formation. *Journal of Cognitive Neuroscience*, *31*(10), 1443–1454. doi: 10.1162/jocn{_}a{_}01411

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1) (No. 1). MIT press Cambridge.

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event Boundaries in Perception Affect Memory Encoding and Updating. *Journal of Experimental Psychology: General*, *138*(2), 236–257. doi: 10.1037/a0015631

Takeuchi, T., Duszkiewicz, A. J., Sonneborn, A., Spooner, P. A., Yamasaki, M., Watanabe, M., … Morris, R. G. M. (2016). Locus coeruleus and dopaminergic consolidation of everyday memory. *Nature*, *537*(7620), 1–18. doi: 10.1038/nature19325

Talmi, D., Lohnas, L. J., & Daw, N. D. (2019). A retrieved context model of the emotional modulation of memory. *Psychological Review*, *126*(4), 455–485. doi: 10.1037/rev0000132

Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005). Adaptive Coding of Reward Value by Dopamine Neurons. *Science*, *307*(5715), 1642–1645. doi: 10.1126/science.1105370

Tulving, E. (1984). Précis of Elements of episodic memory. *Behavioral and Brain Sciences*, *7*(2), 223–238. doi: 10.1017/S0140525X0004440X

Ubl, B., Kuehner, C., Kirsch, P., Ruttorf, M., Diener, C., & Flor, H. (2015). Altered neural reward and loss processing and prediction error signalling in depression. *Social Cognitive and Affective Neuroscience*, *10*(8), 1102–12. doi: 10.1093/scan/nsu158

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.

Villano, W. J., Otto, A. R., Chiemeka, C. E., Gillis, R., & Heller, A. S. (2020). Temporal Dynamics of Real-World Emotion Are More Strongly Linked to Prediction Error Than Outcome. *Journal of Experimental Psychology: General*. doi: 10.1037/xge0000740.supp

Wagatsuma, A., Okuyama, T., Sun, C., Smith, L. M., Abe, K., & Tonegawa, S. (2017). Locus coeruleus input to hippocampal CA3 drives single-trial learning of a novel context. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1714082115

Walters, K., Christakis, D. A., & Wright, D. R. (2018). Are Mechanical Turk worker samples representative of health status and health behaviors in the U.S.? *PLoS ONE*, *13*(6). doi: 10.1371/journal.pone.0198835

Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *15*(4), 263–290. doi: 10.1002/bdm.414

Whitton, A. E., Kakani, P., Foti, D., Van't Veer, A., Haile, A., Crowley, D. J., & Pizzagalli, D. A. (2016). Blunted neural responses to reward in remitted major depression: A high-density event-related potential study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(1), 87–95. doi: 10.1016/j.bpsc

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267–279. doi: 10.1016/j.neuron.2013.11.005

Wimmer, G. E., Braun, E. K., Daw, N. D., & Shohamy, D. (2014). Episodic memory encoding interferes with reward learning and decreases striatal prediction errors. *The Journal of Neuroscience*, *34*(45), 14901–14912. doi: 10.1523/JNEUROSCI.0204-14.2014

Wimmer, G. E., Daw, N. D., & Shohamy, D. (2012). Generalization of value in reinforcement learning by humans. *European Journal of Neuroscience*, *35*(7), 1092–1104. doi: 10.1111/j.1460-9568.2012.08017.x

Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H. J., & Düzel, E. (2005). Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron*, *45*(3), 459–467. doi: 10.1016/j.neuron.2005.01.010

Yang, Y., Cui, Y., Sang, K., Dong, Y., Ni, Z., Ma, S., & Hu, H. (2018). Ketamine blocks bursting in the lateral habenula to rapidly relieve depression. *Nature*, *554*(7692), 317–322. doi: 10.1038/nature25509

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525. doi: 10.1016/j.tins.2011.06.006

Yntema, D. B., & Trask, F. P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, *2*(1), 65–74. doi: 10.1016/S0022-5371(63)80069-9

Young, K. D., Siegle, G. J., Bodurka, J., & Drevets, W. C. (2016). Amygdala activity during autobiographical memory recall in depressed and vulnerable individuals: Association with symptom severity and autobiographical overgenerality. *American Journal of Psychiatry*, *173*(1), 78–89. doi: 10.1176/appi.ajp.2015.15010119

Young, K. D., Siegle, G. J., Zotev, V., Phillips, R., Misaki, M., Yuan, H., ... Bodurka, J. (2017). Randomized clinical trial of real-time fMRI amygdala neurofeedbackfor major depressive disorder: Effects on symptoms and autobiographical memory recall. In *American journal of psychiatry* (Vol. 174, pp. 748–755). doi: 10.1176/appi.ajp.2017.16060637

Yu, A. J., & Dayan, P. (2005). Uncertainty, Neuromodulation, and Attention. *Neuron*, *46*(4), 681–692. doi: 10.1016/j.neuron.2005.04.026

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). *Event perception: A mind-brain perspective* (Vol. 133) (No. 2). doi: 10.1037/0033-2909.133.2.273

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1196.