Behavioral/Cognitive

Episodic Memory Encoding Interferes with Reward Learning and Decreases Striatal Prediction Errors

G. Elliott Wimmer,[1,3] **Erin Kendall Braun,**[1] **Nathaniel D. Daw,**[4] **and Daphna Shohamy**[1,2]
[1]Department of Psychology and [2]Kavli Institute for Brain Science, Columbia University, New York, New York 11027, [3]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg 20246, Germany, and [4]Center for Neural Science and Department of Psychology, New York University, New York, New York 10003

Learning is essential for adaptive decision making. The striatum and its dopaminergic inputs are known to support incremental reward-based learning, while the hippocampus is known to support encoding of single events (episodic memory). Although traditionally studied separately, in even simple experiences, these two types of learning are likely to co-occur and may interact. Here we sought to understand the nature of this interaction by examining how incremental reward learning is related to concurrent episodic memory encoding. During the experiment, human participants made choices between two options (colored squares), each associated with a drifting probability of reward, with the goal of earning as much money as possible. Incidental, trial-unique object pictures, unrelated to the choice, were overlaid on each option. The next day, participants were given a surprise memory test for these pictures. We found that better episodic memory was related to a decreased influence of recent reward experience on choice, both within and across participants. fMRI analyses further revealed that during learning the canonical striatal reward prediction error signal was significantly weaker when episodic memory was stronger. This decrease in reward prediction error signals in the striatum was associated with enhanced functional connectivity between the hippocampus and striatum at the time of choice. Our results suggest a mechanism by which memory encoding may compete for striatal processing and provide insight into how interactions between different forms of learning guide reward-based decision making.

Key words: decision making; hippocampus; learning; memory; reward; striatum

Introduction

Adaptive behavior depends on improving one's decisions by learning from past experience. For example, if we enjoyed a restaurant's sushi many times in the past, we are likely to order it again. We might also recollect the restaurant's location and the friends we ate with each time. In the laboratory, these two forms of learning (reward learning and episodic memory) have each been studied extensively in isolation. In more naturalistic experience, however, they necessarily overlap, raising questions about the nature of such interactions.

Reward learning research has shown how choices are guided by past associations between stimuli and rewards (Daw and Doya, 2006; Schultz, 2006; Rangel et al., 2008). Typically, studies examine how trial-by-trial outcomes update the value of stimuli in simple repeated situations. Reward learning is thought to be driven by reward prediction errors (RPEs), signaled by dopamine neurons projecting to the striatum, a key locus of learning (Barto, 1995; Schultz et al., 1997; Sutton and Barto, 1998; Delgado et al., 2000; Knutson et al., 2000; McClure et al., 2003; O'Doherty et al., 2003; Steinberg et al., 2013).

A separate body of research has focused on the cognitive and neural mechanisms of episodic memory (Eichenbaum and Cohen, 2001). In humans, episodic memory is typically tested by presenting a series of stimuli, each appearing only once, and subsequently testing participants' memory for these episodes (for review, see Paller and Wagner, 2002). Damage to the hippocampus impairs episodic encoding (Eichenbaum and Cohen, 2001) and in fMRI studies successful encoding is associated with activation in the hippocampus (e.g., Wagner et al., 1998; Brewer et al., 1998; Staresina and Davachi, 2009).

How does episodic memory encoding interact with behavioral and neural correlates of reward-based learning? Anatomically, the supporting neural systems are interconnected at multiple levels (Lisman and Grace, 2005; Shohamy and Adcock, 2010; Pennartz et al., 2011; Kahn and Shohamy, 2013). Functionally, multiple studies have demonstrated negative "competitive" correlations between the striatum and the hippocampus during learning (e.g., Packard and McGaugh, 1996; Poldrack et al., 2001), yet thus far there has been a lack of data connecting these negative neural interactions to human behavior. Moreover, emerging data also report positive correlations between activity

Figure 1. Reward learning and memory task. ***A***, In the reward learning task, participants made choices between two options (colored squares), each associated with a drifting probability of reward, with the goal of earning as much money as possible. Reward learning models predict that feedback on trial t (e.g., a reward after choosing green) will influence choice behavior on trial t + 1. Incidental, trial-unique object pictures, unrelated to the reward learning task, were overlaid on each option. Inset, Example win probability sequence for 50 trials. ***B***, A surprise memory test was administered 1 d later. Participants were asked to determine whether the objects were old or new and to rate their confidence.

in these neural systems during decision making and memory (Peters and Büchel, 2010; Sadeh et al., 2011; Scimeca and Badre, 2012; Wimmer and Shohamy, 2012). Thus, many open questions remain about the interactions between episodic memory and reward learning and the implications of these interactions for value-based decision making.

In this study, we examined whether and how encoding long-term episodic memories relates to concurrent reward learning behavior (model parameters and choices) and brain activity (fMRI correlates of prediction error responses in the striatum) by adapting a well-established reward learning task to include incidental episodic encoding. Our novel design provides trial-by-trial measures of reward-guided choices and episodic memory encoding, allowing us to ask questions about interactions between the cognitive and neural systems supporting these types of learning.

Materials and Methods

Participants
A total of 64 subjects participated in the study (32 in the scanner, 32 behaviorally only). Informed consent was obtained in a manner approved by the Columbia University Institutional Review Board. Subjects were fluent English speakers with no self-reported neurological or psychiatric disorders and normal or corrected-to-normal vision. Data from three participants were excluded (one missed >20% of choices in the reward learning task; two had errors in behavioral data collection), leaving 61 subjects (30 scanned, 31 behavioral; 35 female; mean age: 22 years; range: 18–35 years).

The fMRI participants were paid $20 per hour for ~3.5 h of participation, plus one-half of the nominal rewards earned during the experimental task. The behavior-only participants were paid $12 per hour for ~1.5 h of participation, plus one-fifth of the nominal rewards earned during the reward learning task.

For the group analyses, the fMRI and behavioral subjects were combined, as the tasks were identical except for minor procedural differences (noted below), and the combination of the groups allowed for a more powerful and detailed characterization of individual differences.

Procedure
The experiment was conducted across two consecutive days of testing for each subject (Fig. 1). The first day consisted of a reward learning task and the second of a surprise episodic memory test.

Reward learning task (day 1). The reward learning task was based on a classic, well-understood "bandit" reward learning task (e.g., Platt and Glimcher, 1999; Frank et al., 2004; Samejima et al., 2005; Daw et al., 2006; Behrens et al., 2007; Schönberg et al., 2007; Gershman et al., 2009). On each trial, subjects selected one of two options and then received binary monetary reward feedback (Fig. 1A). The two options were distinguished by their color (blue or green). Subjects were instructed that each option was associated with a different probability of reward that would change slowly during the course of the experiment; their goal was to earn the most money in the task by selecting on each trial the option they estimated was most likely to yield a reward.

Choice- and trial-unique object pictures were presented in the center of each option. Subjects were instructed that objects would be presented but that the objects were not relevant to the monetary outcomes. Subjects were familiarized with the reward learning task, including the incidental

objects in a short practice block, which was also repeated inside the scanner.

On each trial, subjects had 2 s to choose between the blue and green options by pressing a corresponding left or right button with their right hand. The left–right location of the blue and green options was permuted randomly from trial to trial. After the 2 s choice period, options remained on the screen for an additional 0.5 s, followed by a blank 1 s interstimulus interval. Then, subjects received "reward" or "miss" feedback for 1.75 s. The $0.25 reward outcome was represented by an image of a quarter dollar, and the $0.00 miss outcome was represented by a phase-scrambled image of a quarter dollar. If no choice was made during the choice period, no feedback was displayed and "Too late!" was presented on the screen during the feedback period. A jittered fixation intertrial interval preceded the next trial (fMRI: mean 3.5 s, range 1.5–13.5 s; behavioral: mean 3 s, range 1.5–9.5 s). To signal trial onset, the fixation cross switched from white to black 1 s before the next trial began.

Each of the two options was associated with a different drifting probability of monetary reward (Fig. 1A, inset). To encourage continual learning, reward probabilities diffused gradually across the 100 task trials according to Gaussian random walks with reflecting boundary conditions at 20% and 80%. Two probability sequences were generated according to this process, each containing a probability sequence for the two options. Two additional sequences were created by reversing the assignment of sequence to options, resulting in four sequences that were counterbalanced across subjects. For each trial and for each subject, observed binary reward and miss outcomes were derived from these underlying reward probability sequences.

Finally, as a control for the presentation of the objects during reward learning, subjects completed another block of reward learning, but without objects being presented. In this control block, shape stimuli represented options (a yellow circle and triangle). This block provided a control for the effect of objects on reward learning and also provided more behavioral and fMRI data for basic reward learning analyses. After the reward learning task, participants were informed of their winnings.

The duration and distribution of intertrial intervals (null events) were optimized for estimation of rapid event-related fMRI responses as calculated using Optseq software (http://surfer.nmr.mgh.harvard.edu/optseq/). The task was presented using MATLAB (MathWorks) and the Psychophysics Toolbox (Brainard, 1997). The reward learning task was projected onto a mirror above the MRI head coil.

Episodic memory test (day 2). Subjects returned to the laboratory and completed a surprise memory test to measure subsequent memory for the objects that had been incidentally presented during the reward learning task (Fig. 1B). The memory test was administered behaviorally. In the memory test, subjects saw each of the 200 "old" objects from the reward learning task (100 trials × 2 objects per trial) intermixed with 100 "new" objects. The order of the old pictures was pseudo-randomized from the reward learning task order, and old and new pictures were pseudo-randomly intermixed. On each trial, a single object was presented above the response options "old" and "new." Subjects pressed the left or right arrow key to indicate whether they thought that the picture was "old" (seen during the reward learning task) or "new" (not previously seen). To avoid any influence of option biases from the reward learning task on subsequent memory responses at test, the objects were presented without the blue and green option colors. Then subjects used the 1–4 number keys to indicate their confidence in the memory response on a rating scale: "guess," "pretty certain," "very certain," and "completely certain." The memory test responses had no time limit and included opportunities for rest breaks to avoid fatigue.

Next, all subjects completed a written questionnaire consisting of questions about choice strategies and attention during both tasks, as well as whether during the reward learning task subjects thought that their memory may be tested for the incidentally presented objects. Finally, subjects were paid for their participation and their reward learning task winnings.

The behavioral subgroup was tested in an experimental testing room. The task was the same as for the fMRI subgroup, except that the behavioral subgroup experienced 80 trials (yielding 160 "old" and 80 "new" images in the memory test), a single pair of reward probability sequences

was used, option color was indicated by a transparent color overlay, and the final control reward learning task (without objects) was not administered.

Behavioral analysis
Analyses of the behavioral data were designed to test for potential interactions between memory and reward learning. First, analyses were conducted to verify, separately, that subjects used reward feedback to guide their choices in the reward learning task and that subjects showed significant memory for the objects presented during reward learning. Next, analyses were conducted to test whether and how episodic memory and reward learning interacted.

Reward learning. Learning was analyzed using computational models, following prior studies (e.g., Daw et al., 2006). To verify learning and examine its form, we used computational models to attempt to explain the series of choices in terms of previous events (for reviews of the methodology, see O'Doherty et al., 2007; Daw, 2011). These included both simple regression models (to test for local adjustments in behavior making minimal assumptions about their form) and a Q-learning reinforcement learning model (which uses a more structured set of assumptions to capture longer-term coupling between events and choices).

In the regression analysis, we used a logistic regression model to account for each subject's sequence of choices in terms of two explanatory variables coding events from the previous trial: the choice made and whether it was rewarded (both coded as binary indicators, implemented in STATA 9.1; StataCorp) (Lau and Glimcher, 2005; Gershman et al., 2009; Daw et al., 2011; Li and Daw, 2011). In a secondary analysis, predictor variables based on choice and reward on three preceding trials were also included.

In the reinforcement learning analysis, we fit a Q-learning reinforcement learning model to subject's choice behavior (Sutton and Barto, 1998). This generated subject-specific model parameters and model fit values (log likelihood). The model also allowed us to compute trial-by-trial value and reward feedback variables for use in behavioral analyses of subsequent memory and neural analyses of parametric fMRI responses.

The reinforcement learning model learns to assign an action value to each option, Q_1 and Q_2, according to previously experienced rewards. These are assumed to be learned by a delta rule: if option c was chosen and reward r (1 or 0) was received, then Q_c is updated according to the following:

$$Q_{c,t} = Q_{c,t-1} + \alpha^*\delta_{c,t} \qquad (1)$$

$$\delta_{c,t} = r_t - Q_{c,t-1} \qquad (2)$$

where the free parameter α controls the learning rate. Option values were initialized with values of 0.5.

Given value estimates on a particular trial, subjects are assumed to choose between the options stochastically with probabilities P_1 and P_2 according to a softmax distribution (Daw et al., 2006):

$$P_{c,t} \propto \exp(\beta(Q_{c,t} + \varphi I(c,c_{t-1}))) \qquad (3)$$

The free parameter β represents the softmax inverse temperature, which controls the exclusivity with which choices are focused on the highest-valued option. The model also included a free parameter φ, which, when multiplied by the indicator function $I(c,c_{t-1})$, defined as 1 if c is the same choice as that made on the previous trial and zero otherwise, captures a tendency to choose (for positive φ) or avoid (for negative φ) the same option chosen on the preceding trial (Lau and Glimcher, 2005; Schönberg et al., 2007). Because the softmax is also the link function for the logistic regression model discussed above, this analysis also has the form of a regression from Q values onto choices (Lau and Glimcher, 2005; Daw, 2011), except here, rather than as linear effects, the past rewards enter via the recursive learning of Q, controlled, in nonlinear fashion, by the learning rate parameters.

Parameters were optimized for each subject using an optimization routine, which included 20 starting points to avoid local minima. To generate per-subject, per-trial values for chosen value and RPE (for subsequent memory and fMRI analyses), the model was simulated on each subject's sequence of experiences. The model was simulated using the

median across-subjects best-fitting learning rate, softmax, and perseveration parameters from the reinforcement learning model. This is because individual parameter fits in tasks and models of this sort tend to be noisy, and regularization of the fit parameters across subjects tends to improve a model's subsequent fit to fMRI data (following previous work: Daw et al., 2006; Schönberg et al., 2007; Wimmer et al., 2012). Additionally, we simulated performance to search for parameters that maximized optimal choices in the reward learning task. Simulations used 100 runs each of 100 different combinations of learning rates (0.01–1) and softmax inverse temperatures (0.1–10.0), with perseveration set at 0.10 and win probability drifts as used in the actual participant population.

Episodic memory. To measure episodic memory, we computed corrected hit rates: the difference between the proportion of hits (old items identified as old) and false alarms (new items incorrectly identified as old). The use of corrected hit rate reduces the influence of response biases that can inflate or reduce raw hit rate. "Guess" responses were removed from this analysis to focus our across-subject analyses on responses that are more consistent with episodic recognition rather than familiarity (following standard procedure in episodic memory studies: e.g., Wagner et al., 1998; Otten et al., 2001).

For within-subject trial-by-trial memory analyses, as memory hit rates were relatively low and because the use of the graded confidence rating scale may differ between subjects, it was necessary to differentiate between confident episodic responses and lower-confidence responses. Thus, we defined an individualized threshold separating lower-confidence responses from higher-confidence responses by examining the reliability of responding at different confidence levels (Parks, 1966).

In particular, we identified for each subject the confidence level above which "old" responses were more reliable than guessing (d′) (Parks, 1966). When d′ is below zero, the proportion of "new" objects that were incorrectly identified as "old" was greater or equal to the proportion of "old" objects that were correctly identified as "old." After excluding "guess" responses, we focused on the first confidence level above guessing ("pretty certain"). For a given subject, if d′ at this level was below zero, we excluded data from this level and examined the next level. If d′ at this level was at or above zero, we then computed d′ across this level and all higher levels of confidence. If this measure was above zero, we included all of these responses. This approach identifies for each subject the lowest confidence threshold above which memory responses, considered as a group of responses all above a given level, are reliable. Using this method, the data from a majority (66%) of subjects included all of the "pretty certain" and above responses and all but one of the remaining subjects included all of the "certain" level and above responses. Conservatively, miss responses were included at all confidence levels.

To ensure the inclusion of sufficient memory events to yield robust estimates in behavioral and fMRI analyses, we excluded subjects with <15 successful higher-confidence memory events in total. This led to the exclusion of 5 subjects (all from the fMRI subgroup), leaving a total of 56 subjects for the analyses of memory and interactions of memory and reward learning.

For trial-by-trial reward learning and memory analyses, we created a single memory success variable. The memory variable was derived from the two potentially remembered stimuli on each trial (the incidental objects presented in chosen and nonchosen options). Memory responses for chosen and nonchosen options were combined (memory for chosen objects related to memory for nonchosen objects: regression analysis, see Results). The combined memory variable also increased our ability to detect any interactions between memory and reward learning. Separately for the chosen and nonchosen options, successfully remembered items of subject-defined higher confidence (as defined above) were coded as 1, remembered items of subject-defined lower confidence were coded as 0, and all forgotten items were coded as −1. Memory values for chosen and nonchosen options were additively combined, yielding a summed memory regressor with values ranging from −2 to 2. In the description of results, we use "remembered" and "forgotten" to refer to the positive and negative values of this memory measure. Trials with positive values of the combined memory measure (which include either one or two higher-confidence remembered objects) will be referred to as "remembered." Trials with negative values of the combined memory measure (which

include either one or two forgotten objects) will be referred to as "forgotten."

Interactions between episodic memory and reward learning: across-participant analyses. Our primary analyses focused on the interaction between memory encoding and reward learning. Across subjects, we tested the correlation between memory performance (corrected hit rate) and the reinforcement learning model learning rate. We also tested for a correlation between memory performance and reinforcement learning model fit (log likelihood). For model fit, as log likelihood is summed across trials, values from the 80 trial behavioral task were normalized to reflect 100 trials. As a supplemental task performance analysis, we tested for a correlation between memory performance and the mean reward probability for the subject's chosen options.

Interactions between episodic memory and reward learning: trial-by-trial within-participant analyses. We used the choice-predicting logistic regression described above, which contains terms for reward and choice (on trial t, t − 1, and t − 2) as predictors of subsequent choice (on trial t + 1). We added to this model an interaction of the summed memory variable and reward (both on trial t). This interaction can be used to test whether reward has a stronger or weaker influence on subsequent choice on trials when incidental objects are successfully remembered.

Additionally, in this regression model, we tested for an interaction between memory on still earlier trials and reward learning. Initial memory analyses indicated that prior trial memory strongly predicted current trial memory (and conversely that prior trial unsuccessful memory predicted unsuccessful memory) ($t_{(53)}$ = 12.53, p < 0.001). Given this correlation, we reasoned that the overall involvement of memory systems may have been changing slowly throughout the task. This fluctuating memory engagement level would be revealed only noisily by memory for objects on a particular trial. To address this possibility, we included predictors for previous trials' memories, which under these conditions (by virtue of being independent samples of the ongoing state of memory engagement) might interact with reward learning, over and above the current trial's memory. The logistic regression model included two additional memory and reward interaction regressors to capture any prior trial memory effects. Specifically, memory on prior trials (trial t − 1 and trial t − 2) was interacted with reward (trial t). If a response was not made on a preceding trial, the prior trial reward, memory, and interaction regressors were set to zero. In a more conservative follow-up analysis, we restricted analysis to trials in which both objects were either remembered or forgotten.

Finally, to ensure that effects were not due to differences between subgroups of subjects (behavioral or fMRI), all regressors were interacted with a group variable, and these interactions were included in the regression analyses.

Control analyses. We further tested whether any interactions between memory and reward learning could be explained by other variables that may relate to attention. To test whether current memory formation influenced current choice, an additional model was estimated that included the influence of memory on the current choice (trial t + 1) interacted with reward (trial t). Next, to examine attentional measures on the prior trial (trial t), separate models were estimated that tested whether control variables interacted with reward and decreased the significance of a memory by reward interaction. Specifically, we tested a binarized measure of fast versus slow choice reaction times (Luce, 1986) and effects of reaction time controlling for choice probability and entropy.

We also examined several other variables that might, analogous to reaction time, serve as markers indicating attentional engagement. In separate models, we tested whether task periods with higher versus lower reward rates over recent trials (quantified by the number of rewards received over the prior five trials) or higher versus lower underlying maximum win probability, interacted with reward or decreased the significance of a memory by reward interaction. Finally, an additional model was estimated that replaced the subject's memory measure in all main effects and interactions with an across-subjects memory measure: each object's mean encoding probability across the group minus the subject's own contribution to the mean.

We also tested the related question of whether memory formation was influenced by reward learning. Using random-effects regression, we tested the correlation between memory and reward. In additional mod-

els, we tested the correlation between memory and RPE or choice value. RPE and choice value were computed from the reinforcement learning model, using the median across-subjects parameter fits.

fMRI data acquisition

Whole-brain imaging was conducted on a 3.0T Phillips MRI system at Columbia University's Program for Imaging and Cognitive Sciences, using a SENSE head coil. Head padding was used to minimize head motion; no subject's motion exceeded 2 mm in any direction from one volume acquisition to the next. Functional images were collected using a gradient echo T2*-weighted EPI sequence with BOLD contrast (TR = 2000 ms, TE = 15 ms, flip angle = 82, $3 \times 3 \times 3$ mm voxel size; 45 contiguous axial slices). For each functional scanning run, five discarded volumes were collected before the first trial to allow for magnetic field equilibration. Four functional runs of 232 TRs (7 min and 44 s) were collected, each including 50 trials. Following the functional runs, structural images were collected using a high-resolution T1-weighted MPRAGE pulse sequence ($1 \times 1 \times 1$ mm voxel size).

fMRI data analysis

Preprocessing and data analysis were performed using AFNI (Cox, 1996) and Statistical Parametric Mapping software (SPM8; Wellcome Department of Imaging Neuroscience). Functional images were coregistered manually using AFNI; the remainder of the analysis was completed in SPM. Images were realigned to correct for subject motion and then spatially normalized to the MNI coordinate space by estimating a warping to template space from each subject's anatomical image and applying the resulting transformation to the EPIs. Images were resampled to 2 mm cubic voxels, smoothed with an 8 mm FWHM Gaussian kernel, and filtered with a 128 s high-pass filter. Data collected on one subject during the 2011 Virginia earthquake were inspected with independent component analysis as implemented in FSL's MELODIC (Beckmann and Smith, 2004) to ensure that no artifacts were introduced by the earthquake.

fMRI model regressors were convolved with the canonical hemodynamic response function and entered into a GLM of each subject's fMRI data. The six scan-to-scan motion parameters produced during realignment were included as additional regressors in the GLM to account for residual effects of subject movement. Linear contrasts of the resulting SPMs were taken to a group-level (random-effects) analysis. We report results corrected for family-wise error (FWE) due to multiple comparisons (Friston et al., 1994). We conduct this correction at the peak level within small-volume ROIs for which we had an *a priori* hypothesis or at the whole-brain cluster level. For planned analyses in the hippocampus, based on a previous report of reward and memory interactions (Adcock et al., 2006), we defined 6 mm radius spherical ROIs centered at (right, anterior, superior: $-20, -10, -18$) and ($20, -12, -18$). In the striatum, peak RPE activation across the reward learning task and control task without objects was used to define bilateral 6 mm radius spheres in the striatum (see Results). These ROIs were combined into hippocampal or striatal ROI masks for small-volume correction. For display purposes, we render all small-volume-corrected (SVC) significant activations at $p < 0.005$. All voxel locations are reported in MNI coordinates, and results are displayed overlaid on the average of all subjects' normalized high-resolution structural images.

To examine interactions between memory and reward learning, mirroring the analyses of the behavioral data, analyses of the fMRI data proceeded in two steps: (1) we tested separately for effects of reward learning and memory encoding; and (2) we conducted several analyses that examined how these effects related to one another.

Reward learning. A GLM tested for BOLD correlates of trial-by-trial reinforcement learning variables. Reinforcement learning model parameters (learning rate, softmax, perseveration) were derived from the median fit across subjects (for rationale, see above). For fMRI analyses, these were derived from a fit across the reinforcement learning task and the control task behavior in the fMRI subgroup. For the reward learning localization analysis only, fMRI models included data from both the task with objects and the control task with no objects. In addition to control regressors during choice and feedback (0 s duration), the model included a parametric regressor (modulating the control regressor) at the time of choice for the value of the chosen option (0 s duration). We used the probability of the chosen action from the softmax equation, which serves to normalize the value against that of the nonchosen option (Daw et al., 2006; Boorman et al., 2009). At the time of feedback, the model included a regressor for RPE (0 s). To provide more statistical power over which to estimate neural correlates of choice value and RPE, we estimated this model over functional data from both the reward learning task and the following control reward learning task without objects.

Additionally, to test whether neural correlates of prediction error expressed both algebraic components of the prediction error signal, reward and choice value, we estimated an additional GLM. In this model, the prediction error regressor was replaced by two parametric regressors at the time of feedback, choice value and reward (Behrens et al., 2008; Li and Daw, 2011; Niv et al., 2012).

Episodic memory. The second GLM tested for correlates of successful memory encoding. In addition to control regressors during the choice and feedback period (2.5 and 0 s duration, respectively), this GLM included a parametric regressor during the choice period that represented the summed memory for chosen and nonchosen items, as described above (2.5 s duration). This model is equivalent to a first-level model with two separate regressors modeling memory for chosen and nonchosen items, where the separate effects are combined additively as a second-level contrast.

Interactions between episodic memory and reward learning. To test how episodic memory interacted with reward learning, our primary GLM analysis tested for neural signals reflecting an interaction of memory and RPE. The interaction was computed by multiplying the summed memory regressor by the RPE regressor. This model included control regressors at choice and feedback (0 s duration) and three parametric regressors during the feedback period: RPE interacted with memory, RPE, and memory. As a control for across-subjects variability in reinforcement learning parameters, we additionally conducted this analysis using individually fit parameters. To visualize the effect of memory on RPE, we estimated an additional GLM where the memory variable only included trials where both objects were remembered or where both objects were forgotten. We extracted β coefficients from this model and computed a contrast for remembered and forgotten trials.

Additionally, to further explore the interaction between memory and prediction error, we tested whether memory related to the reward and choice value components of the prediction error signal. We estimated an additional GLM where the RPE and its interaction with memory, at feedback time, were replaced by four parametric regressors modeling reward, choice value, and both variables' interactions with memory.

Cross-region interactions

To examine memory-related functional connectivity between the hippocampus and the striatum, we performed a psycho-physiological interaction (PPI) analysis (Friston et al., 1997). As a seed region for the PPI analysis, we focused on the hippocampal region showing the strongest response to memory, which overlapped with the *a priori* left hippocampus ROI ($-20, -10, -18$). Signal was extracted from a 6 mm radius sphere around this *a priori* coordinate. The PPI analysis tested for regions showing greater functional correlation with the timecourse of activity in the hippocampus (the physiological variable) depending on memory success (the psychological variable).

To compute the PPI, the timecourse of activation from the hippocampus was extracted and deconvolved. This timecourse was multiplied by the memory indicator (choice period, 2.5 s duration). Because the PPI uses a simple contrast of conditions, we binarized the memory measure, contrasting trials with at least one remembered object versus trials with no remembered objects (of the two objects presented on each choice). We used this memory breakdown because it yielded an approximately equal number of trials in each bin, making it optimal for contrast estimation. This regressor was then convolved with the hemodynamic response function to yield the hippocampus by memory interaction regressor. The PPI GLM included the interaction regressor, the memory regressor, and the unmodulated hippocampus timecourse regressor.

To examine and compare functional connectivity at choice and feedback, two control PPI models were estimated. First, a model was esti-
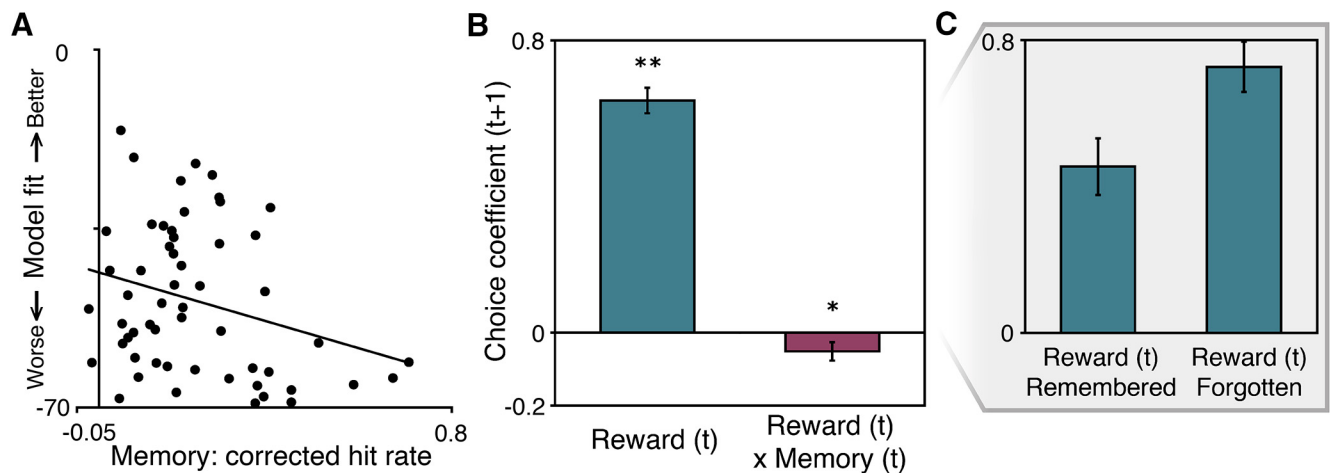
Figure 2. Episodic memory is negatively correlated with reward learning, both across and within participants. *A*, Negative correlation between memory and reinforcement learning model fit (log likelihood) across participants. *B*, Choice prediction by reward and the interaction of memory and reward on a trial-by-trial basis within participants: Reward on the prior choice (trial t) positively predicts subsequent choice (trial t + 1) (blue). The interaction of memory and reward had a negative influence on subsequent choice (purple). This suggests that, on a trial-by-trial basis, better memory is associated with worse reward-based updating. *p < 0.05. **p < 0.01. *C*, Interaction of memory and reward on choice: the net effect of reward on choice is plotted separately for trials with remembered objects and forgotten objects.

mated to test memory-related connectivity during the reward feedback period (1.75 s). Second, to ensure that any differences in connectivity between the choice and feedback periods were not simply due to an increased power to detect effects for the longer-duration choice model (2.5 s vs 1.75 s), a control choice period model was estimated where the choice regressor matched the duration of the feedback regressor (1.75 s).

Results

Reward learning behavioral performance
Participants ($n = 61$) engaged in a two-alternative reward learning task, with the goal to earn as much money as possible. Before turning to our primary analyses testing the interaction between memory and reward learning, we first verified that participants used reward feedback to guide their choices. Using logistic regression, we found that reward for a choice predicted the subsequent choice (0.649 ± 0.034, mean $\pm$ SEM; $t_{(43)} = 19.11$, $p < 0.001$). Similar effects were seen for reward predicting choice on up to two further preceding trials (p values < 0.001), indicating that participants' choices overall were strongly influenced by the history of rewards they received. We also fit a reinforcement learning model to choice behavior (Sutton and Barto, 1998). Mean learning rate was 0.53 ± 0.04, and mean softmax inverse temperature was 7.06 ± 1.62. These mean values are similar to prior reports in this type of task (e.g., Gershman et al., 2009) and are also near optimal parameter values estimated via simulation (learning rate range = 0.60–0.80, softmax range = 4–7).

Episodic memory behavioral performance
The memory test revealed significant subsequent memory for the objects that appeared within choice options, despite the fact that these objects were unrelated to the participants' goal of maximizing their earnings. Over all participants, the mean corrected memory rate, excluding "guess" responses, was $19.70 \pm 2.2\%$ ($t_{(60)} = 8.93$, $p < 0.001$ vs zero). Excluding individuals with poor memory performance (for all analyses going forward; see Materials and Methods), the mean corrected memory rate was $21.8 \pm 2.2\%$ (correct "old" response hit rate: $56.2 \pm 2.1\%$; incorrect "old" response false alarm rate: $34.4 \pm 2.4\%$). The hit rate did not differ between the first and second half of the reward learning task (56.3% vs 56.1%, respectively, $p > 0.96$). The distribution across confidence ratings for "old" objects was as follows: "guess" =

$20.5 \pm 3.2\%$ of items; "pretty certain" = $30.6 \pm 2.7\%$; "very certain" = $24.7 \pm 2.2\%$; "completely certain" = $29.5 \pm 4.2\%$. Memory responses were distributed across the 1, 0, and -1 bins: high-confidence hits: 65.2 ± 3.5 (range 20–120); low-confidence hits: 28.3 ± 3.1 (0–101); misses: 80.4 ± 3.4 (20–142).

Episodic memory negatively interacts with reward learning
Across-participants analyses
To explore potential interactions between memory formation and reward learning, we first tested the relationship between memory performance and reinforcement learning model learning rate and model fit. We found that memory performance (corrected hit rate) was negatively correlated with reinforcement model learning rate ($r = -0.30$, $p < 0.03$). This correlation indicates that participants with better memory performance were slower to update their choices based on reward feedback. Applying a Box-Cox transformation (Box and Cox, 1964) to the learning rate distribution yielded the same result ($r = -0.31$, $p < 0.03$). Memory performance was not significantly related to softmax inverse temperature (a measure of choice "noisiness"; $p > 0.14$).

In addition, we found that memory performance was negatively correlated with reinforcement learning model fit (log likelihood) ($r = -0.36$, $p < 0.006$; Fig. 2A). The correlation with model fit indicates that the behavior of participants with better memory was less well described by a standard reinforcement learning model that predicts choices based on values learned from reward feedback. Similarly, we found a negative correlation between memory performance and a measure of reward learning performance not derived from model fits: the mean reward probability of the participant's chosen options across the task ($r = -0.30$, $p < 0.03$).

Trial-by-trial within-participant analyses
To address the critical question of how memory interacts with reward learning at a trial-by-trial level, within participants, we augmented the logistic regression model of the influence of reward (trial t) on subsequent choice (trial t + 1) with an additional variable capturing the interaction of reward (trial t) and memory encoding success for the objects seen on that trial. We found a significant negative interaction between successful memory en-

coding and the influence of reward on subsequent choice (full regression, $t_{(36)} = -2.06$, $p < 0.04$; Fig. 2B). The same effect was found when restricting the analysis to trials when the objects were either both remembered or both forgotten ($t_{(36)} = -2.31$, $p < 0.03$; Fig. 2C), and this effect did not relate to differential effects for memory for chosen versus nonchosen objects ($p > 0.51$). Importantly, the negative interaction between memory and reward was replicated in a separate behavioral experiment ($n = 20$), where the learning and test sessions were performed on the same day ($t_{(10)} = -2.75$, $p < 0.006$). The finding that reward on trials with successful memory formation had a weaker influence on the subsequent choice suggests that episodic encoding competes with learning from the reward received later on the same trial.

A different possibility is that object encoding competes with the decision task at the level of choice, rather than reward learning. To test this possibility, we examined whether reward interacted with episodic memory for the objects viewed concurrent with the choice (trial t + 1) rather than on the same trial as the reward. However, there was no such interaction ($p > 0.52$), although the prior trial memory effect remained significant ($p < 0.04$). The lack of a memory effect for the object viewed concurrent with choice suggests that the memory–reward interaction is not related to modulations in reward learning task engagement or differential processing (perhaps due to visual attention) of object versus choice stimuli.

Control analyses
Nonetheless, we conducted several control analyses to further investigate whether the negative memory–reward effect was related to measures of attention on the previous trial. Reasoning that reaction time might reflect variations in attention (e.g., Luce, 1986), we tested whether several reaction time proxies for attention either mimicked or explained away the negative memory by reward interaction. We found that reaction time did not interact with reward ($p > 0.61$), whereas the memory–reward effect remained significant ($p < 0.04$). We next removed prominent effects of choice difficulty (entropy) and choice probability from reaction time. We again found no relationship between corrected reaction time and reward ($p > 0.40$), whereas the memory–reward effect remained significant ($p < 0.04$).

In additional control analyses, we tested several reward learning task measures that may also relate to differential attention. We found that better versus worse recent performance in the reward learning task did not interact with reward ($p > 0.57$), nor did periods when the underlying maximum win probability was higher versus lower ($p > 0.96$); the memory–reward effect remained significant in both cases (p values < 0.04). Finally, we replaced each participant's memory values with the across-participant mean memorability of an object, a proxy measure of particularly memorable and salient objects. In this model, we found no significant memory by reward interaction ($p > 0.26$), suggesting that salient objects were not driving the interaction effect. The null results of these control analyses suggest that the memory–reward effect is not related to attention to sensory events but may instead be related to internal selection processes prior to or concurrent with reward feedback on the preceding trial, such as maintaining a working memory trace of the chosen option or "credit assignment" of the outcome to the choice.

Reward learning effects on memory
Having established that subsequent memory performance was associated with differences in reward learning, we next tested whether events during reward learning predicted subsequent

memory. First, we observed a significant negative correlation between reward on a trial and subsequent memory for objects seen on that trial ($t_{(54)} = -2.68$, $p < 0.008$). A similar but weaker negative correlation was observed between memory and RPE ($t_{(54)} = -2.08$, $p < 0.04$). A model including the separate components of RPE (reward and choice value) showed no significant effect of choice value on memory ($p > 0.89$) but a significant effect of reward ($p < 0.009$). The magnitude of the negative correlation between reward and memory was similar across chosen and nonchosen items (chosen: $t_{(54)} = -1.75$, $p < 0.08$; nonchosen: $t_{(54)} = -2.17$, $p < 0.03$). Additionally, we found that memory was enhanced when an object was presented in a chosen versus nonchosen option (chosen hit rate, $68.0 \pm 2.8\%$; nonchosen hit rate, $63.9 \pm 3.1\%$; t test, $t_{(55)} = 3.34$, $p < 0.001$). The memory benefit for chosen options did not come at the cost of memory for nonchosen options on the same trial: memory formation was highly correlated between options ($t_{(54)} = 5.62$, $p < 0.001$), supporting the use of a combined memory measure in the memory and reward learning interaction analyses. Choice reaction time was positively but nonsignificantly related to episodic memory ($p > 0.13$).

Finally, we tested for differences between the behavioral and fMRI subgroups. We found that the direction of the main effects and memory interactions in the choice prediction analysis were the same in both groups but that, when directly compared, the participants in the behavioral group showed a weaker influence of prior reward on choice (full regression, $t_{(36)} = -4.99$, $p < 0.001$). The behavioral subgroup also exhibited better memory performance (26.8% vs 15.7%; t test, $t_{(54)} = 2.68$, $p < 0.01$).

Striatal correlates of RPE and hippocampal correlates of episodic memory
We conducted two initial GLM analyses to localize BOLD responses (henceforth, "activity") correlated with the main effects of reward and successful memory formation. First, we localized regions with activity correlated with a trial-by-trial RPE time-series extracted from computational model fits (using all $n = 30$ fMRI participants). Replicating prior results, we found that RPE correlated with activity in the striatum at feedback (right putamen: $z = 4.62$ ($30, -10, -6$); left putamen: $z = 4.14$ ($-30, -10, 6$); $p < 0.0001$ uncorrected, for ROI localization; Fig. 3A). The RPE correlation peaked in more posterior regions of the ventral putamen than commonly observed (e.g., McClure et al., 2003; Schönberg et al., 2007; however, for similar posterior localization, see O'Doherty et al., 2003; Dickerson et al., 2011). The ventral putamen prediction error correlation was observed both in the reward learning task with pictures and the subsequent control task without pictures; and outside of the putamen, we did not find more anterior striatal correlates, even at a liberal threshold of $p < 0.01$ uncorrected. Further, in a different GLM, we separated the two algebraic components of RPE (reward and choice value) to test whether the striatum positively correlates with reward and negatively correlates with choice value (Behrens et al., 2008; Li et al., 2011; Niv et al., 2012). We found that activity in the right posterior striatum indeed responded to each component (reward: $z = 3.80$ ($30, -12, -6$), $p < 0.001$ uncorrected; chosen option value: $z = -3.21$ ($26, -8, -10$), $p < 0.001$ uncorrected; Fig. 3B).

Next, we localized regions with activity correlated with successful memory formation, focusing on those participants with evidence for significant memory overall ($n = 25$; see Materials and Methods). Replicating prior results, subsequent memory was associated with activation in the bilateral hippocampus during encoding (left: $z = 3.22$ ($-20, -12, -16$), right: $z = 3.08$ (22,
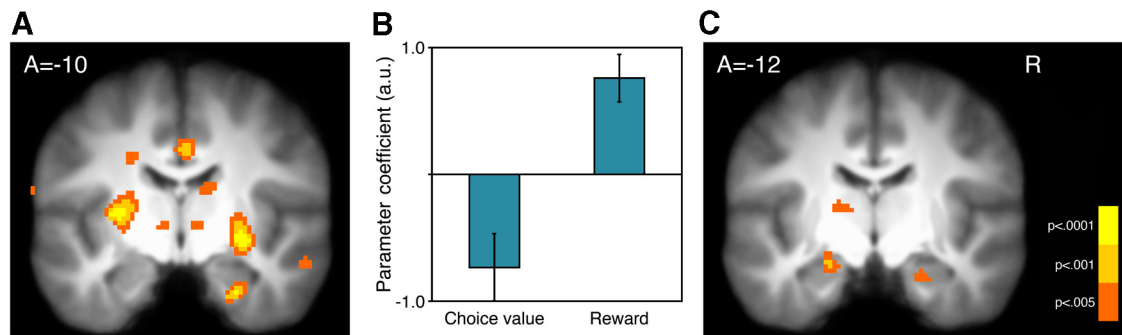
Figure 3. RPE correlates in the striatum and subsequent memory correlates in the hippocampus. **A**, Activation in the striatum during feedback correlated with RPE. **B**, The breakdown of RPE responses into choice value and reward components in the right ventral putamen. **C**, Activation in the hippocampus during choice correlated with subsequent memory (images thresholded at $p <$ 0.005 uncorrected for display).

Table 1. fMRI results

Contrast	Region	Right	Anterior	Superior	z-score	Voxels	p value
RPE	Right putamen	30	−10	−6	4.62	329	<0.001 uncorrected
	Left putamen	−30	−10	−6	4.14	965	<0.001 uncorrected
Memory	Left hippocampus	−20	−12	−16	3.22	47	<0.05*
	Right hippocampus	22	−12	−22	3.08	97	<0.05*
	Right LOC	34	−90	24	5.68	1873	<0.05**
	Left LOC	−34	−88	18	5.2	5521	<0.05**
Memory × reward interaction	Right putamen	26	−6	−4	3.75	40	<0.01*
Memory PPI	Right putamen	34	−2	−6	3.57	85	<0.05*

LOC, Lateral occipital complex.

*SVC FWE-corrected.

**Whole-brain FWE-corrected.
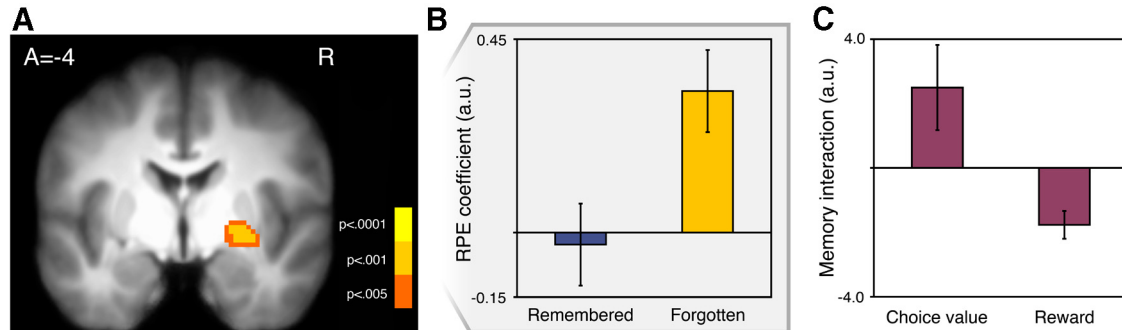
Figure 4. Memory and RPE interaction in the right ventral striatum. **A**, Activation in the striatum correlated with the interaction of memory and RPE. In the right putamen, activation was significantly more correlated with RPE on trials where incidental stimuli were forgotten ($p < 0.05$ SVC; image thresholded at $p < 0.005$ uncorrected for display). **B**, Depiction of the memory and RPE interaction extracted from the right ventral putamen ROI shown in **A** (for display purposes only). **C**, The breakdown of the interaction into the choice value and reward components of RPE in the same ROI.

−12, −22); $p < 0.05$ SVC; Fig. 3C). We additionally found that activity in bilateral object-responsive lateral occipital cortex (Grill-Spector et al., 2001) correlated with subsequent memory ($z = 5.68$ (34, −90, 24); $z = 5.20$ (−34, −88, 18); $p < 0.05$ whole-brain FWE-corrected (Table 1)).

Memory formation negatively interacts with the striatal RPE signal

Behaviorally, we found evidence for negative effects of episodic memory encoding on reward learning; thus, our primary fMRI analysis focused on testing where and when this negative interaction may occur. As memory interacted with the use of reward feedback in guiding subsequent choices, we hypothesized that memory might be associated with a decrease in the striatal response to reward.

To test whether the correlation with RPE in the striatum differed as a function of episodic memory formation, we first examined the interaction of memory and RPE. We indeed found a significant negative interaction of memory and RPE in the striatum ($z = -3.75$ (26, −6, −4), $p < 0.01$ SVC; Fig. 4A). The negative interaction was localized to the region of the striatum that showed the strongest correlation with RPEs during learning, the ventral putamen. The same results held when using individually fit reinforcement learning parameters ($z = 3.49$, $p < 0.05$ SVC). This result indicates that activity was significantly more correlated with RPEs on trials when incidental stimuli were forgotten versus when they were remembered. Indeed, β coefficients extracted from the ventral putamen suggest that the correlation with RPE was reduced to near zero on trials with only remembered objects (Fig. 4B).

A

B
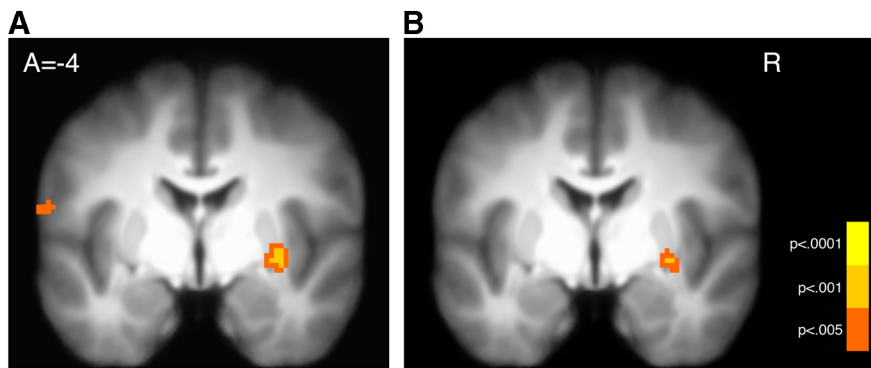
Figure 5. Hippocampal–striatal connectivity associated with memory. **A**, Activity in the hippocampus is significantly more correlated with activity in the striatum during trials with remembered versus forgotten objects ($p < 0.05$ SVC). **B**, A conjunction of the memory PPI (**A**) and the interaction of memory and RPE (Fig. 4A) illustrates overlapping effects in the ventral striatum (images thresholded at $p < 0.005$ uncorrected for display; conjunction threshold: $p < 0.005$).

To examine this interaction in more detail, we conducted a separate GLM analysis where RPE was replaced with its reward and choice value components. We found that memory interacted with each variable in the expected direction in the striatum: memory by reward, $z = -3.14$ (26, −6, −8), $p < 0.05$ SVC; memory by choice value, $z = 3.00$ (26, −8, −2), $p < 0.05$ SVC (Fig. 4C). The separate effects on reward and choice value suggest that memory encoding success related both to the sensory processing of reward feedback as well as to the internal representation of choice value. Overall, the negative interaction of memory and RPE shows that on trials with successfully remembered stimuli the striatal response corresponded less closely to RPEs drawn from the Q-learning model, whereas responses during choice were unaffected. This difference in feedback processing could underlie our behavioral finding of a decreased influence of reward on choice when incidental stimuli are successfully encoded.

Finally, we examined the effect of memory on choice value at the time of choice instead of at feedback. Although both the right ventral putamen ($p < 0.05$, SVC) and the ventromedial PFC ($p < 0.0001$, uncorrected) positively correlated with choice value, we found that memory encoding was not related to the choice value correlation in these regions (even at a relaxed threshold of $p < 0.01$, uncorrected).

Memory formation increases hippocampal–striatal connectivity

To explore the negative interaction between memory and RPE, we tested whether hippocampal–striatal connectivity was related to successful memory formation. We conducted a PPI testing for a greater correlation between hippocampal activation and striatal activation on trials with remembered versus forgotten objects focusing on the region of the left hippocampus, which exhibited the strongest memory correlation, and the region of the right ventral striatum negatively correlated with the interaction of memory and RPE.

We found that better memory was associated with an increase in correlation between hippocampal and striatal activation (right ventral putamen: $z = 3.67$ (34, −4, −6), $p < 0.05$ SVC; Fig. 5). This increase in connectivity occurred at the time of choice, but not later in the trial at the time of reward feedback. At feedback, we found no memory-related difference in hippocampal–striatal correlations, even at a relaxed statistical threshold ($p < 0.01$ uncorrected). At choice, no regions outside the striatum survived whole-brain correction, but a more anterior region of the ventral

striatum showed a similar effect as the right putamen ($z = 4.75$ (8, 4, −2), $p < 0.0001$ uncorrected). Our finding that hippocampal–striatal connectivity was positively related to memory suggests that the interfering effects of memory formation on RPE signaling may originate in hippocampal memory encoding processes.

Discussion

The present results demonstrate interactions between reward learning and episodic memory. We found a negative interaction between reward-based updating of choices and memory for details about the choice options. These results indicate that episodic memory encoding can trade off with the use of trial-by-trial outcomes to guide choices during learning. This trade-off was evident in behavior, both within and across participants, as well as in the representation of RPEs in the striatum. These results suggest that, when receiving rewards, better episodic memory for the events preceding the reward may be associated with worse reward learning.

Our findings advance our understanding of memory and reward learning in several ways. First, our results increase the understanding of how seemingly separate forms of learning influence each other and work together to guide behavior. The demonstration that episodic memory and reward learning do not operate in isolation is broadly consistent with an emerging recognition of the importance of interactions between multiple cognitive and neural processes in both learning and decision making (Doya, 1999; Poldrack and Packard, 2003; Daw et al., 2005; Pennartz et al., 2011; Delgado and Dickerson, 2012; Shohamy and Turk-Browne, 2013), extending the traditional focus in systems and cognitive neuroscience on mapping distinct cognitive processes onto independent and distinct neural regions (Knowlton et al., 1996; Squire and Zola, 1996).

Second, our results demonstrate a trade-off between these two forms of learning, reflected both in trial-by-trial behavior and in activity in the underlying neural systems. Although there has been increasing interest in the nature of the interaction between multiple learning systems, findings have been inconsistent, leading to debates about whether the interaction between different systems is best characterized as cooperation or competition (Packard and McGaugh, 1996; Poldrack and Packard, 2003; Burgess, 2006; Pennartz et al., 2011). Lesions of either system have been shown to enhance the type of performance supported by the remaining system (e.g., Lee et al., 2008). Additionally, animal studies have also shown that, over the course of extensive learning, behavioral control shifts from the hippocampus to the striatum (Packard and McGaugh, 1996). Our results are consistent with both the cooperation and competition frameworks. In support of competition, we show evidence for a negative interaction between forms of memory classically related to the striatum and the hippocampus. Further, our results build upon previous across-task studies suggesting a competitive relationship by providing novel trial-by-trial measures of this interaction, providing novel evidence for such a relationship at both the behavioral and neural levels in a single study.

At the same time, a simple competition framework does not provide a full account for the interactions between these systems even within our experiment: we also show that successful

memory formation is related to greater functional connectivity between the hippocampus and the striatum during choice, consistent with other reports of cooperation between these systems (Sadeh et al., 2011). Although we find no direct relationship between the two effects, on successful encoding trials with greater hippocampal–striatal connectivity, we also find reduced feedback responses in the striatum. As a caveat, it is difficult to separate BOLD responses in a rapid event-related design; however, our results depend not on separating choice and feedback responses but on decomposing the feedback response by memory. Overall, our results add to a growing body of research demonstrating that the nature of the interaction between different forms of learning is not unitary or simple and is likely to depend on the demands of the environment.

Indeed, although it is tempting to interpret our results in terms of a competition between episodic memory and reward learning, where better performance in one comes at the cost of worse performance in the other, another possibility is that reward learning in the presence of successful episodic memory is different, but not actually worse. In particular, the episodic memory system is well situated to guide choices (Lengyel and Dayan, 2005; Biele et al., 2009), although memory-guided choices likely reflect different quantitative principles than standard, incremental reinforcement learning models. According to this view, choices engaging the episodic memory system might be accompanied by effects similar to what we observe in our data, including better episodic encoding, enhanced hippocampal–striatal interactions, and choices and neural prediction errors that are more poorly fit by standard reinforcement learning models. Speculatively, the particular pattern of the behavioral effect, whereby learning from reward trades off with episodic memory, might be due to credit assignment of the received reward to the trial-unique object rather than the option color. A direct test of this idea would be possible in a task where objects occur more than once.

Although our results focus on the effect of successful memory encoding, a number of lines of evidence suggest that decisions may be guided by retrieval of non-procedural memories as well as (or instead of) incremental trial-and-error reward learning (Tolman, 1948; Dickinson, 1985; Daw et al., 2005; Lengyel and Dayan, 2005; Daw and Shohamy, 2008; Biele et al., 2009). For example, it has been suggested that hippocampal memories support choices by simulating candidate future courses of action (Johnson and Redish, 2007; Buckner, 2010; Pfeiffer and Foster, 2013) or new decisions (Wimmer and Shohamy, 2012; Barron et al., 2013), and that decisions may be guided by episodic memory (Lengyel and Dayan, 2005; Biele et al., 2009) or working memory (Collins and Frank, 2012) for previous choices.

Given this menagerie of models, at present there is much less clarity about the trial-by-trial, quantitative progression of these sorts of learning, compared with the classic and quantitatively specific incremental reward learning theories that drove our analysis. Our results may thus reflect the possibility that such novel influences contribute more to choices when episodic memory is strong. Indeed, a slower learning rate and, equivalently, a weaker modulation of choice by an immediately preceding reward is consistent with a longer timescale for hippocampal learning (Bornstein and Daw, 2012). Neurally, many such forms of learning are less likely to rely on RPEs, at least in their standard form (Gläscher et al., 2010; Daw et al., 2011). Future studies that affirmatively characterize the contributions of alternative models to incremental reward learning will be needed to test this possibility.

A different candidate explanation for many of our effects might be due to attentional fluctuations between the objects and the reward learning task. One version of this idea is that some of our effects are due to competition for sensory processing of different task cues, as by visual attention. Importantly, though, our effects were specific, allowing us to reject the possibility that our results relate trivially to fluctuations in task engagement, instead isolating specific aspects of the learning behavior that are subject to competition. For instance, if enhanced object memory was due simply to sensory attention to the object cues at the expense of the reward learning task, we would have expected to see reaction times (a classic correlate of on-task attention) (Luce, 1986) driving the effect, but they do not. Moreover, we did not find evidence that successful memory encoding competed with the use of values to guide the choice on the same trial, as expected under direct attentional competition between the object and task: the negative influence of memory encoding was selectively related to a weaker effect of reward on subsequent choice. This suggests competition at the level of learning values from reward, rather than choice. Indeed, the competition between object memory and learning from reward cannot be due to direct competition for sensory processing between object and reward cues because they are not displayed simultaneously. Further evidence that the effects of memory extend beyond simple attentional processing of sensory cues is that the effect of memory on striatal RPEs comprises not just the reward itself but also the component of prediction error related to choice value, a subjective, internal variable. It is also possible that attention is allocated sequentially on a trial, first to the value-guided choice, and then to the objects, in such a way that successful encoding interferes with the maintenance of working memory for which option is chosen. In other incidental encoding tasks, however, neural correlates of encoding success appear before 400 ms (Sanquist et al., 1980; Paller et al., 1987), faster than choices in our task (mean 867 ms). Further, we find no relationship between reaction time and the effect of reward on choice, as might have been expected if the effect were driven by competition between the representations during the time remaining after the choice.

In any case, our results together suggest that competition must act at the level of internal representations, namely, the memory for the chosen option or in the subsequent differential assignment of feedback to objects versus cues. These results may reflect the brain's mechanisms for key computational features of reinforcement learning: the maintenance of "traces" of chosen options as well as the "credit assignment" of outcomes to preceding choices. Although such mechanisms are not yet well understood in neuroscience or psychology, they are likely related to cue competition phenomena in learning (e.g., Mackintosh, 1975; Pearce and Hall, 1980; Robbins, 1998; Kruschke, 2001). Such competition also includes as a special case another topic of ongoing interest, the control of the rate of learning even to a single cue (Dayan et al., 2000; Behrens et al., 2007; Li et al., 2011).

Finally, from the perspective of what influences episodic memory, our results demonstrate that reward might impair memory formation. Superficially, this result seems to contrast with recent studies reporting that reward can enhance episodic memory (Wittmann et al., 2005; Adcock et al., 2006; Bialleck et al., 2011; Mather and Schoeke, 2011). A critical difference may be that prior experiments used static reward contingencies and did not include choice and learning. A further difference is that, in prior studies, reward value was already known to participants when objects were presented (Wittmann et al., 2005; Adcock et al., 2006). Further studies will be necessary to more fully under-

stand the positive and negative effects of reward learning on memory.

In conclusion, research in episodic memory and reward learning has progressed tremendously in the past few decades, yet these areas remain predominantly independent. Our results suggest that an important goal for future research is to further understand the interrelationships between the cognitive and neural systems supporting memory and reward learning. Greater knowledge of these interactions is important because these two types of learning often co-occur. For example, outside of simple experimental settings, rewards and punishments are usually associated with episodes in life that contain unique items and contexts. Our results demonstrate that such experiences can be influenced by multiple interactive forms of learning, providing new insight into how learning can guide decisions and actions.

References

Adcock RA, Thangavel A, Whitfield-Gabrieli S, Knutson B, Gabrieli JD (2006) Reward-motivated learning: mesolimbic activation precedes memory formation. Neuron 50:507–517. CrossRef Medline

Barron HC, Dolan RJ, Behrens TE (2013) Online evaluation of novel choices by simultaneous representation of multiple memories. Nat Neurosci 16:1492–1498. CrossRef Medline

Barto AG (1995) Adaptive critics and the basal ganglia. In: Models of information processing in the basal ganglia (Davis JL, Houk JC, Beiser DG, eds), pp 215–232. Cambridge, MA: Massachusetts Institute of Technology.

Beckmann CF, Smith SM (2004) Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imaging 23:137–152. CrossRef Medline

Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. Nat Neurosci 10:1214–1221. CrossRef Medline

Behrens TE, Hunt LT, Woolrich MW, Rushworth MF (2008) Associative learning of social value. Nature 456:245–249. CrossRef Medline

Bialleck KA, Schaal HP, Kranz TA, Fell J, Elger CE, Axmacher N (2011) Ventromedial prefrontal cortex activation is associated with memory formation for predictable rewards. PLoS One 6:e16695. CrossRef Medline

Biele G, Erev I, Ert E (2009) Learning, risk attitudes and hot stoves in restless bandit problems. J Math Psychol 53:155–167. CrossRef

Boorman ED, Behrens TE, Woolrich MW, Rushworth MF (2009) How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. Neuron 62:733–743. CrossRef Medline

Bornstein AM, Daw ND (2012) Dissociating hippocampal and striatal contributions to sequential prediction learning. Eur J Neurosci 35:1011–1023. CrossRef Medline

Box GEP, Cox DR (1964) An analysis of transformations. J R Stat Soc Ser B 26:211–252.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436. CrossRef Medline

Brewer JB, Zhao Z, Desmond JE, Glover GH, Gabrieli JD (1998) Making memories: brain activity that predicts how well visual experience will be remembered. Science 281:1185–1187. CrossRef Medline

Buckner RL (2010) The role of the hippocampus in prediction and imagination. Annu Rev Psychol 61:27–48, C1–C8. CrossRef Medline

Burgess N (2006) Spatial memory: how egocentric and allocentric combine. Trends Cogn Sci 10:551–557. CrossRef Medline

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

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173. CrossRef Medline

Daw ND (2011) Trial-by-trial data analysis using computational models In: Attention and performance, Vol XXIII (Delgado MR, Phelps EA, Robbins TW, eds), pp 3–38. Oxford: Oxford UP.

Daw ND, Doya K (2006) The computational neurobiology of learning and reward. Curr Opin Neurobiol 16:199–204. CrossRef Medline

Daw ND, Shohamy D (2008) The cognitive neuroscience of motivation and learning. Soc Cogn 26:593–620. CrossRef

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704–1711. CrossRef Medline

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

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

Dayan P, Kakade S, Montague PR (2000) Learning and selective attention. Nat Neurosci 3 [Suppl]:1218–1223.

Delgado MR, Dickerson KC (2012) Reward-related learning via multiple memory systems. Biol Psychiatry 72:134–141. CrossRef Medline

Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA (2000) Tracking the hemodynamic responses to reward and punishment in the striatum. J Neurophysiol 84:3072–3077. Medline

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

Dickinson A (1985) Actions and habits: the development of behavioural autonomy. Philos Trans R Soc Lond B 308:67–78. CrossRef

Doya K (1999) What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? Neural Netw 12:961–974. CrossRef Medline

Eichenbaum H, Cohen NJ (2001) From conditioning to conscious recollection: memory systems of the brain. New York: Oxford UP.

Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306:1940–1943. CrossRef Medline

Friston KJ, Worsley KJ, Frackowiak RS, Mazziotta JC, Evans AC (1994) Assessing the significance of focal activations using their spatial extent. Hum Brain Mapp 1:210–220. CrossRef Medline

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218–229. CrossRef Medline

Gershman SJ, Pesaran B, Daw ND (2009) Human reinforcement learning subdivides structured action spaces by learning effector-specific values. J Neurosci 29:13524–13531. CrossRef Medline

Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66:585–595. CrossRef Medline

Grill-Spector K, Kourtzi Z, Kanwisher N (2001) The lateral occipital complex and its role in object recognition. Vision Res 41:1409–1422. CrossRef Medline

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

Kahn I, Shohamy D (2013) Intrinsic connectivity between the hippocampus, nucleus accumbens, and ventral tegmental area in humans. Hippocampus 23:187–192. CrossRef Medline

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

Knutson B, Westdorp A, Kaiser E, Hommer D (2000) fMRI visualization of brain activity during a monetary incentive delay task. Neuroimage 12:20–27. CrossRef Medline

Kruschke JK (2001) Toward a unified model of attention in associative learning. J Math Psychol 45:812–863. CrossRef

Lau B, Glimcher PW (2005) Dynamic response-by-response models of matching behavior in rhesus monkeys. J Exp Anal Behav 84:555–579. CrossRef Medline

Lee AS, Duman RS, Pittenger C (2008) A double dissociation revealing bidirectional competition between striatum and hippocampus during learning. Proc Natl Acad Sci U S A 105:17163–17168. CrossRef Medline

Lengyel M, Dayan P (2005) Hippocampal contributions to control: the third way. In: Advances in neural information processing systems, Vol 20 (Platt J, Koller D, Singer Y, Roweis S, eds), pp 889–896. Cambridge, MA: Massachusetts Institute of Technology.

Li J, Daw ND (2011) Signals in human striatum are appropriate for policy update rather than value prediction. J Neurosci 31:5504–5511. CrossRef Medline

Li J, Schiller D, Schoenbaum G, Phelps EA, Daw ND (2011) Differential roles of human striatum and amygdala in associative learning. Nat Neurosci 14:1250–1252. CrossRef Medline

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

Luce RD (1986) Response times: their role in inferring elementary mental organization. New York: Oxford UP.

Mackintosh NJ (1975) A theory of attention: variations in the associability of stimuli with reinforcement. Psychol Rev 82:276–298. CrossRef

Mather M, Schoeke A (2011) Positive outcomes enhance incidental learning for both younger and older adults. Front Neurosci 5:129. CrossRef Medline

McClure SM, Berns GS, Montague PR (2003) Temporal prediction errors in a passive learning task activate human striatum. Neuron 38:339–346. CrossRef Medline

Niv Y, Edlund JA, Dayan P, O'Doherty JP (2012) Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. J Neurosci 32:551–562. CrossRef Medline

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

O'Doherty JP, Hampton A, Kim H (2007) Model-based fMRI and its application to reward learning and decision making. Ann N Y Acad Sci 1104:35–53. CrossRef Medline

Otten LJ, Henson RN, Rugg MD (2001) Depth of processing effects on neural correlates of memory encoding: relationship between findings from across- and within-task comparisons. Brain 124:399–412. CrossRef Medline

Packard MG, McGaugh JL (1996) Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. Neurobiol Learn Mem 65:65–72. CrossRef Medline

Paller KA, Wagner AD (2002) Observing the transformation of experience into memory. Trends Cogn Sci 6:93–102. CrossRef Medline

Paller KA, Kutas M, Mayes AR (1987) Neural correlates of encoding in an incidental learning paradigm. Electroencephalogr Clin Neurophysiol 67:360–371. CrossRef Medline

Parks TE (1966) Signal-detectability theory of recognition-memory performance. Psychol Rev 73:44–58. CrossRef Medline

Pearce JM, Hall G (1980) A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol Rev 87:532–552. CrossRef Medline

Pennartz CM, Ito R, Verschure PF, Battaglia FP, Robbins TW (2011) The hippocampal–striatal axis in learning, prediction and goal-directed behavior. Trends Neurosci 34:548–559. CrossRef Medline

Peters J, Büchel C (2010) Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal–mediotemporal interactions. Neuron 66:138–148. CrossRef Medline

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

Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. Nature 400:233–238. CrossRef Medline

Poldrack RA, Packard MG (2003) Competition among multiple memory systems: converging evidence from animal and human brain studies. Neuropsychologia 41:245–251. CrossRef Medline

Poldrack RA, Clark J, Paré-Blagoev EJ, Shohamy D, Creso Moyano J, Myers C, Gluck MA (2001) Interactive memory systems in the human brain. Nature 414:546–550. CrossRef Medline

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

Robbins TW (1998) Arousal and attention: psychopharmacological and neuropsychological studies in experimental animals. In: The attentive brain (Parasuraman R, ed), pp 189–220. Cambridge, MA: Massachusetts Institute of Technology.

Sadeh T, Shohamy D, Levy DR, Reggev N, Maril A (2011) Cooperation between the hippocampus and the striatum during episodic encoding. J Cogn Neurosci 23:1597–1608. CrossRef Medline

Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. Science 310:1337–1340. CrossRef Medline

Sanquist TF, Rohrbaugh JW, Syndulko K, Lindsley DB (1980) Electrocortical signs of levels of processing: perceptual analysis and recognition memory. Psychophysiology 17:568–576. CrossRef Medline

Schönberg T, Daw ND, Joel D, O'Doherty JP (2007) Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. J Neurosci 27:12860–12867. CrossRef Medline

Schultz W (2006) Behavioral theories and the neurophysiology of reward. Annu Rev Psychol 57:87–115. CrossRef Medline

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

Scimeca JM, Badre D (2012) Striatal contributions to declarative memory retrieval. Neuron 75:380–392. CrossRef Medline

Shohamy D, Adcock RA (2010) Dopamine and adaptive memory. Trends Cogn Sci 14:464–472. CrossRef Medline

Shohamy D, Turk-Browne NB (2013) Mechanisms for widespread hippocampal involvement in cognition. J Exp Psychol Gen 142:1159–1170. CrossRef Medline

Squire LR, Zola SM (1996) Structure and function of declarative and nondeclarative memory systems. Proc Natl Acad Sci U S A 93:13515–13522. CrossRef Medline

Staresina BP, Davachi L (2009) Mind the gap: binding experiences across space and time in the human hippocampus. Neuron 63:267–276. CrossRef Medline

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

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: Massachusetts Institute of Technology.

Tolman EC (1948) Cognitive maps in rats and men. Psychol Rev 55:189–208. CrossRef Medline

Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL (1998) Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. Science 281:1188–1191. CrossRef Medline

Wimmer GE, Shohamy D (2012) Preference by association: how memory mechanisms in the hippocampus bias decisions. Science 338:270–273. CrossRef Medline

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

Wittmann BC, Schott BH, Guderian S, Frey JU, Heinze HJ, Düzel E (2005) Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. Neuron 45:459–467. CrossRef Medline