Behavioral/Systems/Cognitive

Feedback Timing Modulates Brain Systems for Learning in Humans

Karin Foerde and Daphna Shohamy

Department of Psychology, Columbia University, New York, New York 10027

The ability to learn from the consequences of actions—no matter when those consequences take place—is central to adaptive behavior. Despite major advances in understanding how immediate feedback drives learning, it remains unknown precisely how the brain learns from delayed feedback. Here, we present converging evidence from neuropsychology and neuroimaging for distinct roles for the striatum and the hippocampus in learning, depending on whether feedback is immediate or delayed. We show that individuals with striatal dysfunction due to Parkinson's disease are impaired at learning when feedback is immediate, but not when feedback is delayed by a few seconds. Using functional imaging (fMRI) combined with computational model-derived analyses, we further demonstrate that healthy individuals show activation in the striatum during learning from immediate feedback and activation in the hippocampus during learning from delayed feedback. Additionally, later episodic memory for delayed feedback events was enhanced, suggesting that engaging distinct neural systems during learning had consequences for the representation of what was learned. Together, these findings provide direct evidence from humans that striatal systems are necessary for learning from immediate feedback and that delaying feedback leads to a shift in learning from the striatum to the hippocampus. The results provide a link between learning impairments in Parkinson's disease and evidence from single-unit recordings demonstrating that the timing of reinforcement modulates activity of midbrain dopamine neurons. Collectively, these findings indicate that relatively small changes in the circumstances under which information is learned can shift learning from one brain system to another.

Introduction

Learning from the outcomes of actions is central to adaptive behavior. In everyday life, outcomes are sometimes immediate, but are often delayed by seconds, hours, or even days. Despite major advances in understanding the neural mechanisms that support learning from immediate outcomes (Schultz, 1998; O'Doherty et al., 2004), it remains unknown whether learning from delayed outcomes depends on the same or different neural systems.

Research examining learning from immediate feedback has established an essential role for the striatum and its dopaminergic inputs (Schultz, 1998; Pessiglione et al., 2006). However, recent electrophysiological data show that, when rewards are delayed briefly, responses in dopaminergic neurons are fundamentally changed (Fiorillo et al., 2008; Kobayashi and Schultz, 2008), indicating that this mechanism is not well suited for learning from

delayed feedback. Thus, the role of dopamine and the striatum in reward-driven learning may be limited to situations in which rewards arrive immediately following a cue or a response. If so, this raises the question of how learning is accomplished in the many situations in which feedback is not immediate.

We hypothesized that the hippocampus could play an essential role in learning from feedback that is delayed. This proposal is guided by the observation that the hippocampus supports relational learning that binds disparate elements of experiences across space or time (Cohen and Eichenbaum, 1993; Thompson and Kim, 1996; Shohamy and Wagner, 2008; Staresina and Davachi, 2009). Thus, the hippocampus is well suited to support learning from delayed feedback and could complement the role of the striatum in learning from immediate feedback.

Here, we used converging methods to address the following question: Does the timing of feedback have consequences for the cognitive and neural processes supporting learning? To determine the causal role of nigrostriatal mechanisms in learning from immediate versus delayed feedback, Experiment 1 examined learning in patients with Parkinson's disease, which is characterized by dramatic loss of nigrostriatal dopaminergic neurons even in the earliest stages (Agid et al., 1989). Parkinson's disease leads to deficits in incremental feedback-driven learning (Frank et al., 2004; Shohamy et al., 2004), but prior investigations have been limited to situations that involve learning from immediate, response-contingent feedback. Here, we tested the prediction that Parkinson's disease leads to a selective impairment in learning from immediate feedback, but not from delayed feedback.

To examine the dynamic roles of multiple brain systems in learning from immediate versus delayed feedback, in Experiment 2 we used fMRI combined with computational reinforcement-learning models in healthy participants. We predicted that learning from immediate feedback would engage the striatum, whereas learning from delayed feedback would engage the hippocampus.

Finally, we tested whether learning would differ qualitatively as a consequence of feedback timing by including a test of episodic memory for feedback images in our design. In humans, the hippocampus is known to support long-term memory for episodes or events (Davachi, 2006), and based on evidence from animal studies, it has been suggested that learning that depends on the hippocampus (but not the striatum) may result in better memory for feedback events (White and McDonald, 2002).

Materials and Methods

Experiment 1: learning in Parkinson's disease

Participants

Twenty-two participants with a diagnosis of idiopathic Parkinson's disease were recruited from the Center for Parkinson's Disease and Other Movement Disorders at the Columbia University Medical Center Department of Neurology with the assistance of Dr. Lucien Cote. Patients were in mild to moderate stages of the disease (Hoehn and Yahr stages 1–3). Controls, matched on age and education, were recruited from the community surrounding Columbia University. A group of young controls recruited from Columbia University were also tested for comparison with Experiment 2 but were not part of the main analyses in Experiment 1. All participants provided informed consent in accordance with the guidelines of the Institutional Review Board of Columbia University and were paid $12/h for their participation. Participants were excluded if they had suffered brain injury, been diagnosed with neurological or psychiatric disorders other than Parkinson's disease, or if they were on antidepressants or medications affecting the cholinergic system. Participants completed a series of neuropsychological tests and were excluded if they exhibited any general cognitive impairment (scoring 27 or below on the Mini-Mental State Exam) or signs of depression [scoring 7 or above—>2 SDs above the mean for age-matched controls—on the Beck Depression Inventory (BDI) cognitive subscale].

The remaining 18 Parkinson's disease patients and 25 control participants did not differ in age, education, or measures of IQ and frontal executive functions (all values of $p > 0.05$) (Table 1). Of the Parkinson's patients, 13 were being treated with L-Dopa and dopamine agonists and were tested while on their standard medication, 4 were not receiving dopaminergic medications, and 1 did not complete their regular dose before testing; these subgroups were too small to evaluate separately, but analyses restricted to patients on standard dopamine medication ($N = 13$) replicated the results obtained for the whole group: selective impairment for immediate feedback learning (Immediate: $t_{(36)} = -2.66$, $p = 0.01$; Delay: $t_{(36)} = -0.55$, $p = 0.58$).

Task

Participants engaged in a probabilistic learning task similar to tasks previously shown to be sensitive to striatal function in humans (Knowlton et al., 1996; Poldrack et al., 2001; Shohamy et al., 2004; Foerde et al., 2006). Such tasks require participants to learn to associate cues with outcomes through trial and error. Because the relationship between cues and outcomes is probabilistic, there is no one-to-one mapping between cues and outcomes. Thus, optimal learning depends on participants' use of response-contingent feedback to incrementally learn the most probable outcome across multiple trials.

We manipulated the timing with which feedback was delivered in the probabilistic learning task. As illustrated in Figure 1, participants saw a cue (one of four different butterflies) on each trial and had to predict which of two outcomes (differently colored flowers) that cue was associated with. Each butterfly was associated with one flower on 83% of trials and with the other flower on 17% of trials (Table 2). For each butterfly,

Table 1. Demographic characteristics of participants in Experiment 1

	Parkinson's patients	Healthy controls
Age	63.8 (7.8)	65.7 (8.3)
Education	16.1 (1.7)	16.2 (1.6)
MMSE	29.5 (0.8)	29.5 (0.6)
COWAT	44.1 (9.7)	44.6 (11.3)
NAART	17.3 (5.2)	15.8 (8.5)
Digit span	11.9 (1.5)	12.0 (1.9)
UPDRS	31.7 (15.5)	—
Years since onset	7.6 (6.9)	—

Table shows mean (SD). MMSE, Mini-Mental State Examination; COWAT, Controlled Oral Word Association Test; NAART, North American Adult Reading Test; UPDRS, Unified Parkinson's Disease Rating Scale.

Groups did not differ on any measures (values of $p > 0.5$).
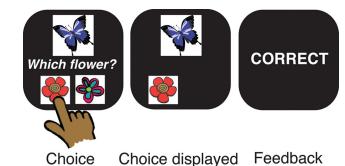

Figure 1. Paradigm for probabilistic learning. Participants used trial-by-trial feedback to learn which flower four different butterflies preferred. On each trial, as soon as a response was made, participants' choices were displayed along with the butterfly until feedback was provided.

feedback followed after a fixed delay of 0 s (Immediate condition) or 6 s (Delay condition), such that two butterflies were associated with each

Table 2. Task structure for Experiment 1

Stimulus Butterfly[b]	Association with outcome		Time between response and feedback[a]
	Flower 1	Flower 2	Delay (s)
1	0.83	0.17	1
2	0.17	0.83	1
3	0.83	0.17	7
4	0.17	0.83	7

[a]After a response was made, the participant was shown their choice for 1 s in all conditions. The delay time shown includes the 1 s choice display plus 0 or 6 s delay time.

[b]Butterflies were counterbalanced with regard to delay condition across participants. For each individual participant, the mapping between cues and delays was fixed.

delay condition (Fig. 2; Table 2). Feedback consisted of the word "CORRECT" or "INCORRECT" displayed for 2 s. The assignment of cues to outcomes and conditions was counterbalanced across participants. Immediate and delayed feedback trial types were interleaved throughout training (Fig. 2).

Participants had up to 7 s to make a response and were given a reminder to respond after 4 s. After responding, they were immediately shown their choice for 1 s followed by the delay period (0 vs 6 s). The chosen flower and the butterfly remained on the screen during the delay to minimize working memory demands. Thus, the critical manipulation was the time interval between response and feedback. Because response times could vary across trials and participants, the overall trial length (butterfly onset to feedback end) could vary, but the time between responses and feedback was always held constant for each trial type (Table 2).

To ensure that participants understood the task and were able to respond in the allotted time they completed a short practice. Next they completed 96 learning trials of the task (Learning phase). Finally, there
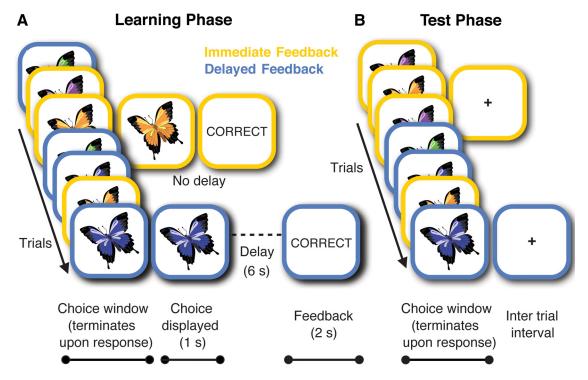
Figure 2. Feedback timing for probabilistic learning with immediate versus delayed feedback. Participants used trial-by-trial feedback to learn which flower four different butterflies preferred (Learning phase). For one set of butterflies (outlined in orange), feedback was presented immediately. For another set of butterflies (outlined in blue), feedback was presented with a delay. After learning, participants completed a probe test in which they continued to make predictions about the butterflies' preferences (Test phase). However, they no longer received feedback, and the timing of all trial events was equal across trial types.

was a Test phase where participants saw the butterflies from the Learning phase and were told to continue performing based on what they had learned. The Test phase resembled the Learning phase, with the exception that no feedback was given and the timing of all trial parts was equivalent across trial types (Fig. 2).

Experiment 2: fMRI of learning in young, healthy individuals
In Experiment 2, we used fMRI combined with computational reinforcement-learning models to examine the dynamic roles of multiple brain systems in learning from immediate versus delayed feedback.

Participants
Data from 20 adults, recruited from the Columbia University campus, are reported (mean age, 23.4 ± 4.1; seven females). All provided informed consent in accordance with the guidelines of the Institutional Review Board of Columbia University and were paid $20/h for their participation. All were right-handed and were screened for pregnancy, use of drugs or psychopharmacological medication, history of neurological damage, and fMRI contraindications. Three additional participants were excluded: one for infrequent responding (responded on 70% of trials across the experiment) and two due to image artifacts or brain abnormalities. A separate group of 25 participants, recruited from the Columbia University campus and paid $12/h for their participation, completed a nonscanned version of the experiment. Task materials and procedures were identical with the scanned study.

Task
Experiment 2 used a parallel version of the task in Experiment 1. The task was modified to adjust the difficulty level to the younger population and to accommodate the task for fMRI. Participants engaged in a probabilistic learning task that required learning to associate six cues (Asian characters) with two different outcomes ("A" or "B") through trial and error. Each character was associated with one outcome on 80% of trials and with the other outcome on 20% of trials (Table 3). For each character, feedback always followed with a fixed delay: 0 s (Immediate condition), 3 s (Short delay condition), or 6 s (Delay condition), such that two

Table 3. Task structure for Experiment 2

Stimulus Character[b]	Association with outcome		Time between response and feedback[a]
	Category A	Category B	Delay (s)
1	0.8	0.2	1
2	0.2	0.8	1
3	0.8	0.2	4
4	0.2	0.8	4
5	0.8	0.2	7
6	0.2	0.8	7

[a]After a response was made, the participant was shown their choice for 1 s in all conditions. The delay time shown includes the 1 s choice display plus 0, 3, or 6 s delay time.

[b]Characters were counterbalanced with regard to delay condition across participants. For each individual participant, the mapping between cues and delays was fixed.

characters were associated with each delay condition (counterbalanced across participants). Trial types for each feedback delay condition were interleaved throughout training.

Participants had up to 3 s to make a response. After responding they were immediately shown their choice for 1 s, followed by the delay period (0, 3, or 6 s). The chosen outcome and character remained on the screen during the delay to minimize working memory demands. Thus, the critical manipulation was the time interval between responses and feedback. Because response times could vary across trials and participants, the overall trial length (character onset to feedback end) could vary, but the time between responses and feedback was constant for each trial type (Table 3). After the delay, performance feedback was displayed for 1.5 s. Feedback was provided in the form of a photograph of an outdoor (correct) or indoor (incorrect) scene.

Before scanning, participants completed a short practice to ensure that they understood that outdoor and indoor scenes signified correct and incorrect responses, respectively. Participants completed 180 learning trials across six runs (trial types were equally distributed across all six blocks) of fMRI scanning (Learning phase). After the Learning phase followed a Test phase in which participants were shown the characters from the Learning phase and were told to continue performing based on

what they had learned. The test resembled the Learning phase, with the exception that no feedback was given and the timing of all trial parts was identical across trial types.

Stimulus presentation sequence and timing were optimized using the optseq2 algorithm (http://surfer.nmr.mgh.harvard.edu/optseq/). Each learning run lasted 374 s. Across all six learning runs, the mean intertrial interval (ITI) was 3.6 s, median was 2.5 s, and range was 0.5–15.5 s. The final probe test run lasted 214 s; mean ITI was 2.27 s and range was 0.5–12.5 s.

Once outside the scanner, ~30 min after completing the probabilistic learning task, participants were given a surprise memory test for the feedback images (indoor and outdoor scenes) they saw during the learning phase. Each image shown during the Learning phase (targets) and an equal number of new images (foils) were presented on a Macintosh PowerbookG4. On each trial, a single image was presented and participants were instructed to determine whether the image was seen during learning (Old) or not seen (New). They were then required to indicate their level of confidence in their choice, with 1, certain; 2, sure; 3, pretty sure; and 4, guessing. The proportion of indoor versus outdoor scenes was equal for target and foil images. Therefore, a strategy of assuming that either outdoor or indoor images were more likely to be targets or foils could not aid performance. Subsequent memory data from one participant were lost due to computer malfunction.

Data analyses

Probabilistic learning. Performance on the probabilistic learning task was assessed both in terms of making optimal choices (the degree to which participants selected the most likely outcome for each cue), and in terms of matching the actual outcome on each trial. The effects of delay and block on these two performance scores were tested in repeated-measures ANOVAs with Huynh–Feldt correction for nonsphericity when appropriate.

Model-derived analyses of learning. To assess the role of feedback in driving learning, we used computational reinforcement-learning models, an approach which has been used extensively in recent studies of reward prediction. Model-derived estimates successfully capture behavior in studies in which participants make choices based on expectations of monetary gain or primary reward. These estimates are designed to index response parameters that are not directly observable in choice behavior (Sutton and Barto, 1998; Daw and Doya, 2006; Pessiglione et al., 2006; Schönberg et al., 2007; Daw, 2011). We estimated trial-by-trial errors in prediction of feedback and then used these estimates in the functional neuroimaging data analysis to test (1) whether the neural responses to feedback were modulated as predicted by reinforcement learning models and (2) whether this effect differed across feedback timing conditions.

We estimated four parameters, learning rates for each of the three feedback delays and a β term (softmax inverse temperature), to optimize the likelihood of the observed behavioral data. The learning rate estimates indicate how sharply the model-predicted outcome expectation for each choice option was updated toward the actual feedback received. β is an index of choice randomness (i.e., the degree to which choices are directed toward the action currently thought to have the highest value), with larger β values indicating less random choice patterns.

On each trial, the predicted value V for choices was updated according to $V = V + \text{lr}^* (\text{outcome} - V)$. Here, the outcome could be 1 (correct) or 0 (incorrect). V was initially set to 0.5. The model estimated the likelihood of each subject's observed choices of A or B for each of the six cues across learning. A separate learning rate was estimated for cues associated with each delay. The probability of participants' choices was computed according to a softmax rule. The optimal sets of parameters for each individual were determined using the maximum log likelihood (Schönberg et al., 2007; Daw, 2011).

The group-averaged parameters were used to apply the fit model to each participant's learning data and create trial-by-trial estimates of feedback prediction errors (Schönberg et al., 2007; Daw, 2011). These feedback prediction errors were then used as parametric regressors at the time of each feedback event in the analysis of the neuroimaging data. There was no linear effect on learning rate across Feedback Timing conditions ($F_{(1,19)} = 2.40$; $p = 0.14$). Therefore, we used the average learning rate across Feedback Timing conditions to generate prediction error regressors.

Subsequent memory for feedback events. To determine whether later memory for feedback events differed for immediate versus delayed feedback, we calculated the proportion of Hits (recognizing previously seen images of indoor and outdoor scenes) that had been associated with each delay during learning and the proportion of False Alarms (incorrectly identifying a new image as previously seen). A corrected hit rate (Hits minus False Alarms) was calculated separately for outdoor and indoor images (because they belonged to distinct categories present in unequal numbers), and the average corrected hit rate across image categories was computed. Performance was further binned according to confidence ratings. Ratings of 1 or 2 were considered high confidence, and these responses were the focus of the subsequent memory analyses.

fMRI methods

Whole-brain imaging was conducted on a 1.5 T GE Signa Twin Speed Excite HD MRI system (GE Healthcare). Structural images were collected using a T2-weighted flow-compensated spin-echo pulse sequence [TR, 3 s; TE, 70 ms; 24 contiguous 5-mm-thick slices parallel to the anterior commissure (AC)–posterior commissure (PC) plane]. Functional images were collected using a T2*-weighted two-dimensional gradient echo spiral-in/out pulse sequence (TR, 2 s; TE, 40; flip, 84; FOV, 22.4; 64 × 64 voxels; 24 contiguous 4.5-mm-thick slices parallel to the AC–PC plane). Spiral in/out has been shown to enhance signal in areas that are vulnerable to susceptibility artifacts, such as the medial temporal lobe, one of our regions of interest (Glover and Law, 2001). The first three volumes from each run were discarded.

Preprocessing and statistical analysis of fMRI data were performed using SPM2 (Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm/). Functional images were corrected for differences in slice acquisition time and for head motion. Individuals' functional and anatomical data were coregistered, normalized to a standard T1 template image, and smoothed with a Gaussian kernel (8 mm full-width half-maximum).

fMRI analyses

Data were analyzed within the framework of the general linear model. In general, trial events were modeled as impulses convolved with a canonical hemodynamic response function and its first-order temporal derivative. Motion parameters were included as covariates of no interest. The SPM2 small volume correction (SVC) procedure was implemented using the familywise error (FWE). A priori anatomical regions of interest (ROIs) were generated by combining the Harvard–Oxford Probabilistic Atlas' (FSL; provided by the Harvard Center for Morphometric Analysis) putamen, caudate, and nucleus accumbens bilaterally to form a striatal ROI, and the hippocampus, parahippocampal gyrus and amygdala bilaterally to form a medial temporal lobe ROI (thresholded at a 25% probability of being in each structure). The anterior hippocampus ROIs were the portions of the hippocampal ROIs anterior to $y = -23$ (the dividing line between the anterior and posterior parahippocampal gyrus ROI). Nonanatomical ROIs were 6 mm spheres centered on voxels of peak activation. All resulting contrast maps were overlaid on a mean anatomical image.

Because the learning task involved a manipulation of feedback time between conditions, it required that feedback always be delivered with the same delay for the cues within each condition. As a result, the event timing of the experimental design necessitated elimination of the temporal variability between responses and feedback that would allow optimal estimation of the BOLD response, in particular for immediate feedback. Thus, direct contrasts between delay conditions were not the basis of the main analyses and interpretations. Instead, we performed analyses by collapsing across delays or making comparisons within delay conditions. Two basic approaches were taken in analyzing the fMRI data: (1) model-derived prediction error analysis and (2) ROI time course analysis.

Model-derived prediction error analysis. The prediction error analysis used the reinforcement learning model-generated estimates of feedback prediction errors (collapsed across all three feedback timing conditions) as parametric regressors at the time of feedback delivery and also mod-
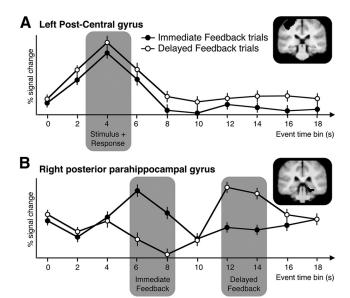
Figure 3. Estimation of event timing in control brain regions outside the learning-related regions of interest. To compare feedback responses across conditions in ROI analyses, time courses were extracted from independent, anatomically defined "control" regions and the critical time points associated with the Immediate versus Delay feedback conditions were estimated. Time courses extracted from regions demonstrating stimulus and feedback task events are plotted. ***A***, Time points illustrating the STIMULUS plus RESPONSE event of the task were extracted from the left post-central gyrus. ***B***, Time points illustrating immediate and delayed FEEDBACK events were extracted from the right parahippocampal gyrus, widely known to respond to images of scenes (Epstein and Kanwisher, 1998). The percentage signal change across time bins at 6 and 8 s were averaged for the Immediate feedback condition and time bins at 12 and 14 s for the Delayed feedback condition. Error bars represent ±1 SEM.

eled trial onsets. ROIs identified in this analysis were then interrogated for effects of feedback timing, which was not included as a factor in the parametric prediction error analysis.

ROI time course analysis. To extract time courses from ROIs, we completed analyses that modeled the trial onsets for correct and incorrect feedback trials, separately for each delay condition resulting in six condition regressors. Deconvolution of signal for each feedback timing condition within ROIs was done using a finite impulse response function implemented with MarsBar (http://marsbar.sourceforge.net/).

We assessed the critical time points associated with trial stimulus onset across conditions and feedback delivery separately for each feedback delay condition independently from our a priori ROIs. As demonstrated in Figure 3, time courses extracted from control ROIs illustrate the feasibility of identifying reliable event times within trials. We determined that the 4 s time bin captured the response to trial stimulus onset and response across conditions (Fig. 3*A*). To determine the time bins for feedback delivery, we extracted time courses from the parahippocampal gyrus because it consistently responds to indoor and outdoor scenes (Epstein and Kanwisher, 1998). The 6–8 s time bin captured the response to immediate feedback and the 12–14 s time bin captured the response to the feedback that was delayed by an additional 6 s (Fig. 3*B*). The intermediate feedback delay condition was omitted from all ROI time course analyses to avoid using overlapping bins between conditions; the response for the intermediate delay condition occurred at the 10 s time bin with no consistent additional time bin. Percentage signal change was then averaged for conditions of interest from the identified time points and analyzed in repeated-measures ANOVAs.

Results
Experiment 1
We assessed the percentage of optimal responses made in the postlearning Test phase and compared performance of Parkinson's disease patients and age-matched controls in the Immediate versus Delayed feedback conditions, as shown in Figure 4. An
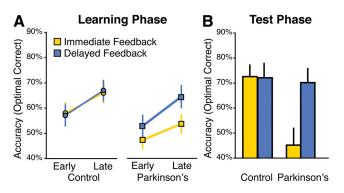

Figure 4. Learning from immediate versus delayed feedback in Parkinson's disease (Experiment 1). Parkinson's patients were selectively impaired after learning from immediate feedback but performed as well as controls when learning from delayed feedback. ***A***, Learning phase. ***B***, Test phase. Error bars represent ±1 SEM.

ANOVA revealed a significant interaction between Feedback Timing and Group ($F_{(1,41)} = 4.7$; $p = 0.036$) as well as main effects of Feedback Timing ($F_{(1,19)} = 4.39$; $p = 0.043$) and Group ($F_{(1,41)} = 6.56$; $p = 0.014$). The interaction was driven by the selective impairment in learning from immediate feedback in the Parkinson's patients (Immediate, $t_{(41)} = -3.43$, $p = 0.001$; Delay, $t_{(41)} = -0.23$, $p = 0.82$). Notably, in the Test phase timing was equal for all trial types and no feedback was given (Fig. 2).

This pattern of selective impairment for immediate feedback was also present during the Learning phase: for Immediate feedback conditions, the group difference was marginally significant early in learning (in the first half of learning trials) ($t_{(41)} = -1.94$; $p = 0.059$) and was significant late in learning (in the second half of trials) ($t_{(41)} = -2.26$; $p = 0.029$). For the Delayed feedback condition, there were no group differences either early or late in learning (all values of $t < 1$) (Fig. 4*A*). There were no significant differences in response times between conditions or groups during Learning or Test phases (all values of $p > 0.05$), and no measures of cognitive function or disease severity were correlated with the impairment in learning from immediate feedback for Parkinson's patients.

Thus, striatal dysfunction was associated with a selective impairment in learning from immediate feedback paired with intact performance when feedback was delayed. In contrast, age-matched controls exhibited no differences in learning as a function of feedback delay. These results suggest that previously reported learning deficits in Parkinson's disease (Shohamy et al., 2004) are selective to learning that is driven by immediate feedback, and that these deficits can be remediated by prolonging the delay between a response and feedback by several seconds. The results also suggest that when feedback is delayed, learning may shift from the nigrostriatal system impaired in Parkinson's disease to alternative neural systems that are spared. In Experiment 2, we used fMRI in healthy individuals to examine this shift and to identify the brain systems that underlie learning from delayed feedback.

Experiment 2
Experiment 2 used a modified version of the task used in Experiment 1 to adjust the difficulty level for the younger population and to accommodate the task for fMRI. As shown in Figure 5, performance accuracy improved over the course of the task and, as in the older healthy controls in Experiment 1, accuracy did not differ as a function of Feedback Timing (Fig. 5*A*; main effect of Block, $F_{(5,95)} = 26.21$, $p < 0.001$; no main effect of Feedback
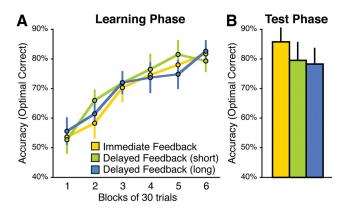
Figure 5. Behavioral performance among young healthy participants in the fMRI study (Experiment 2). Young healthy participants showed equivalent levels of performance across all feedback conditions. **A**, Learning phase. **B**, Test phase. Error bars represent ±1 SEM.

Table 4. Reinforcement model parameters: model used for fMRI analyses

		Mean	(SEM)
Learning rate $(\alpha)^a$	IMMED (1s)	0.22	(0.06)
	DELAY (4s)	0.49	(0.09)
	DELAY (7s)	0.34	(0.08)
	Across conditions	0.35	(0.06)
Softmax inverse temperature (β)	Across conditions	5.10	(1.39)
Pseudo-r^2		0.27	(0.04)
Model fit (log likelihood)		93.22	(4.68)
Chance model fit		124.77	
Likelihood ratio test (model vs chance)		$p < 0.0001$	
Softmax inverse temperature (β) (α fixed to 1)		0.77	(0.07)
Model fit (α fixed to 1)		106.89	(2.80)

^{a}Learning rates did not differ as a function of delay (linear effect: $F_{(1,19)} = 2.40$, $p = 0.14$).

Timing, $F_{(2,38)} = 0.17$, $p = 0.85$, and no Block by Feedback Timing interaction, $F_{(10,190)} = 0.54$, $p = 0.84$). In the Test phase, performance did not differ as a function of Feedback Timing ($F_{(2,38)} = 1.2$; $p = 0.31$; Fig. 5B). Response times did not differ across conditions in either phase (main effect of Block, $F_{(5,85)} = 9.93$, $p < 0.0001$; no significant effect of Feedback Timing, $F_{(2,34)} = 0.16$, $p = 0.85$; and no significant Block by Feedback Timing interaction, $F_{(10,170)} = 0.95$, $p = 0.49$).

Model fitting
We also examined learning rates estimated from a standard reinforcement learning model (see Materials and Methods). As with performance accuracy, learning rates estimated from the reinforcement learning model did not differ as a function of Feedback Timing (linear effect, $F_{(1,19)} = 2.40$, $p = 0.14$). To assess the success of the model in capturing subject behavior, we compared our model to the nested dummy model using a likelihood ratio test (Daw, 2011) and found that our model performed significantly better than chance ($p < 0.0001$) (Tables 4, 5). We also fit a model that estimated separate fits for each delay condition, but found no linear effects for fit ($F_{(1,19)} = 0.22$; $p = 0.65$) or learning rates ($F_{(1,19)} = 0.315$; $p = 0.58$). Therefore, we used the average of learning rates across conditions estimated from the simpler model in the subsequent model-derived fMRI analyses.

These results show that young, healthy participants were able to learn from feedback both when it was immediate and when it was delayed. As seen in previous studies in healthy participants, different cognitive and neural mechanisms may support similar learning performance (Poldrack et al., 2001; Foerde et al., 2006). Thus, our next step was to test whether different neural systems

were engaged in support of learning from immediate versus delayed feedback.

Feedback prediction errors correlate with activation in the hippocampus and the ventral striatum
To explore the relationship between neural activity and participant responses, we looked for areas of the brain in which changes in BOLD correlated with model-derived estimates of feedback prediction errors on a trial-by-trial basis. We collapsed across feedback timing conditions and used a single average learning rate to generate parametric regressors that expressed the error in prediction at the time of feedback delivery (Pessiglione et al., 2006; Schönberg et al., 2007; Daw, 2011). The resulting activation maps are shown in Figure 6.

Notably, this analysis revealed that feedback prediction error estimates were correlated with BOLD activation in bilateral hippocampus (Fig. 6A, Table 6). Additionally, activation in ventral and dorsal striatum (Fig. 6B,C) was correlated with prediction error estimates, consistent with findings from studies of reward learning (Schultz, 1998; O'Doherty et al., 2004; Schönberg et al., 2007).

ROI analyses of immediate versus delayed feedback
Our main prediction was that the hippocampus supports feedback-driven learning when bridging a temporal gap—that is, when feedback is delayed. To address this central question, we conducted a set of analyses comparing feedback processing ROIs, identified in the prediction-error analysis described above, within the ventral striatum (customarily the primary focus of prediction error analyses) and the hippocampus. We estimated feedback sensitivity by comparing responses to correct versus incorrect feedback broken down by Feedback Timing (immediate vs delay; omitting the intermediate condition) (see Materials and Methods).

Consistent with our hypothesis, the hippocampus was selectively sensitive to delayed feedback, but not to immediate feedback (Fig. 6D). A repeated-measures ANOVA revealed a significant three-way interaction [Region (left ventral striatum vs left hippocampus) by Feedback Timing (immediate vs delayed) by Feedback Outcome (correct vs incorrect), $F_{(1,19)} = 4.91$, $p = 0.04$]. Further analyses indicated that responses in the hippocampus were significantly greater to correct than incorrect feedback only when feedback was delayed ($t_{(19)} = 2.62$; $p = 0.017$) but not when it was immediate ($t_{(19)} = 0.22$; $p = 0.83$). For the immediate condition, we found significantly greater response to correct than incorrect feedback in the ventral striatum ($t_{(19)} = 3.07$; $p = 0.006$) (Fig. 6E). This difference between the ventral striatum and the hippocampus was also observed when comparing right ventral striatum and right hippocampus. Thus, the hippocampus was engaged in feedback-driven learning specifically when feedback was delayed.

Feedback prediction errors correlate with activation in the dorsal striatum
We also found that BOLD activity was correlated with feedback prediction errors in left ($p_{SVC\_FWE} = 0.07$) and right ($p_{SVC\_FWE} = 0.003$) dorsal striatum collapsed across delay conditions (Fig. 6C, Table 6). To understand whether timing affected feedback sensitivity in the dorsal striatum, we again compared feedback responses as a function of Feedback Timing in the dorsal striatum ROIs to responses in the hippocampus. A three-way ANOVA revealed interactions between Feedback Timing and Feedback Outcome: $F_{(1,19)} = 5.07$, $p = 0.036$ for left dorsal striatum and left hippocampus (Fig. 6), and $F_{(1,19)} = 5.35$, $p = 0.032$ for right

Table 5. Reinforcement model parameters: model estimating separate fits for each feedback timing condition

	IMMED (1 s)		DELAY (4 s)		DELAY (7 s)	
	Mean	(SEM)	Mean	(SEM)	Mean	(SEM)
Learning rate (α)[a]	0.16	(0.05)	0.33	(0.06)	0.21	(0.06)
Softmax inverse temperature (β)	2.56	(0.37)	2.49	(0.37)	2.60	(0.41)
Pseudo-r^2	0.24	(0.04)	0.29	(0.05)	0.22	(0.04)
Model fit (log likelihood)[b]	31.54	(1.77)	29.33	(2.07)	32.41	(1.63)
Chance model fit	41.59		41.59		41.59	
Likelihood ratio test	$p = 0.005$		$p < 0.0005$		$p = 0.003$	

[a]Learning rates did not differ as a function of delay (linear effect: $F_{(1,19)} = 0.315$, $p = 0.58$).

[b]Fit did not differ as a function of delay (linear effect: $F_{(1,19)} = 0.22$, $p = 0.65$).
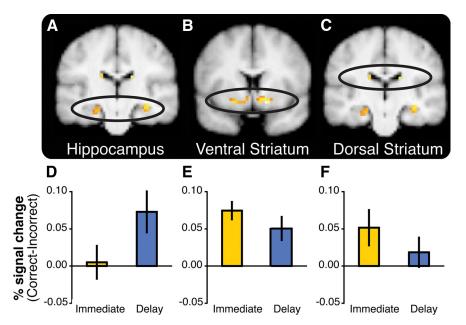

Figure 6. Activation in the hippocampus and the striatum is differentially sensitive to feedback timing during learning. *A–C*, BOLD activity in the hippocampus and the striatum correlated with model-derived feedback prediction errors, collapsed across Immediate and Delayed feedback conditions during learning ($p_{SV\_FWE} < 0.05$; maps displayed at $p < 0.001$ uncorrected within a mask consisting of the striatum and the medial temporal lobe). *D–F*, ROI analyses revealed that the hippocampus was selectively sensitive to delayed feedback, whereas the striatum was sensitive to immediate feedback. The bar graphs represent feedback sensitivity (BOLD percentage signal change difference between Correct and Incorrect trials) extracted from ROIs (6 mm spheres). *D*, Left hippocampus [−24 −18 −18]. *E*, Left ventral striatum [−21 6 −12]. *F*, Left dorsal striatum [−15 −12 21]. Error bars represent ±1 SEM. Parallel results were obtained in anatomical ROIs.

dorsal striatum and right hippocampus. These effects were driven by significantly greater responses to correct than incorrect feedback in both the left ($t_{(19)} = 2.20$; $p = 0.04$) and right ($t_{(19)} = 2.95$; $p = 0.008$) dorsal striatum for the immediate feedback condition. Only the right dorsal striatum showed feedback sensitivity for the delayed feedback condition (right, $t_{(19)} = 2.58$, $p = 0.018$; left, $t_{(19)} = 0.87$, $p = 0.39$). These results are consistent with the idea that immediate feedback may drive stimulus–response learning mechanisms to a greater degree than does delayed feedback and that this is the mechanism that is particularly impaired in Parkinson's disease.

Feedback prediction error analysis of immediate versus delayed feedback
The unique pattern of activation in the hippocampus in response to delayed feedback prompted us to return to investigate whether feedback prediction errors would differ as a function of delay within the hippocampus and the striatum. A direct comparison of prediction errors for immediate versus delayed feedback revealed greater activation in the hippocampus correlated with delayed feedback prediction errors, as demonstrated in Figure 7. In

contrast, immediate prediction errors were correlated with greater activation in the dorsal striatum, consistent with prior studies of feedback-driven learning (Schultz, 1998; O'Doherty et al., 2004; Schönberg et al., 2007).

The difficulty of estimating the BOLD response to feedback separate from the stimulus in the immediate condition created the risk of overweighting activation in the delayed feedback condition. Nonetheless, directly comparing prediction errors between immediate and delayed feedback did not yield a globally (undifferentiated) increased response for the delay condition. Instead, viewing the results from this analysis at a low threshold corroborated the patterns of activation described above and, importantly, suggested that this pattern was not restricted to a few choice voxels but instead represented a region-specific pattern (Fig. 7). This assertion was also supported by ROI analyses using anatomical ROIs from the Harvard–Oxford Probabilistic Atlas. These analyses showed the same pattern obtained when using peaks from the prediction error analysis: Only anterior hippocampus, not caudate or nucleus accumbens, showed selective feedback sensitivity for delayed feedback.

In summary, the fMRI data revealed that activation in the hippocampus was correlated with model-estimated prediction errors during feedback-driven learning. Specifically, the hippocampus was engaged selectively when feedback was delayed, but not when it was immediate, whereas the ventral and dorsal striatum were engaged for immediate feedback. These results are consistent with the findings from Parkinson's patients in Experiment 1, which indicated an essential and selective role for nigrostriatal circuitry in learning driven by immediate feedback. Together, these converging findings reveal complementary roles for the hippocampus and the striatum in feedback-driven learning depending on feedback timing.

Episodic memory for feedback events
Finally, although there were no differences in probabilistic learning as a function of feedback timing in young healthy adults, we wanted to assess other behavioral markers that would indicate whether distinct learning systems were engaged as a function of feedback timing. We hypothesized that engagement of the hippocampus during learning would lead to better episodic memory

Table 6. Cluster maxima from prediction error analysis

Location	~BA	MNI coordinates x	y	z	Z score	FWE p^a
Striatum						
Right ventral putamen		15	9	−12	4.89	<0.001
Left posterior putamen		−30	−12	−9	4.85	0.001
Right caudate body		18	−9	24	4.5	0.003
Left ventral putamen		−21	6	−12	4.29	0.008
Left caudate body		−15	−12	21	3.62	0.072
Left posterior caudate body		−18	−21	24	3.6	0.077
Medial temporal lobe						
Right hippocampus		36	−21	−12	4.06	0.02
Left hippocampus		−24	−18	−18	3.9	0.034
Right hippocampus		27	−9	−21	3.79	0.048
Outside ROIs^b						
Right post-central gyrus	4	24	−30	63	5.97	<0.001
Left middle temporal gyrus	37	−54	−63	0	5.37	0.002
Left superior parietal	40	−33	−48	57	5.09	0.01
Left amygdala		−24	0	−15	5.08	0.01

^{a}FWE $p < 0.05$, whole brain; extent, 5 voxels. Coordinates of cluster maxima locations from feedback prediction error analysis of fMRI data are reported within our ROIs (the striatum and the medial temporal lobe) using SVC and outside our ROIs corrected for multiple comparisons across the whole brain.

^{b}None of these regions showed selective sensitivity for immediate or delayed feedback.
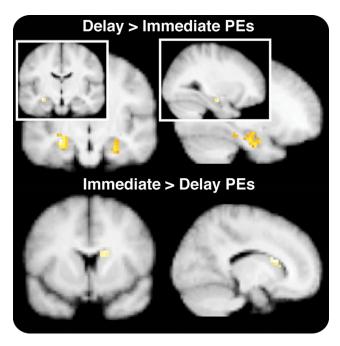

Figure 7. Feedback prediction errors for immediate and delayed feedback conditions. Model-derived prediction errors were estimated separately for each delay. Viewing these direct contrasts at a threshold of $p < 0.05$ restricted to the hippocampus and the striatum revealed a greater correlation between prediction errors and BOLD activity for Delayed than Immediate feedback in the hippocampus. This effect was also apparent at a slightly more conservative threshold of $p < 0.005$ (see inset). In contrast to the hippocampus [−30 −12 −18], the dorsal striatum [15 15 15] showed a greater correlation with prediction errors for immediate than for delayed feedback ($p < 0.05$).

for feedback events. This prediction about memory for feedback events themselves follows from the well known role for the hippocampus in supporting long-term memory for episodes or events (often referred to as episodic memory; for review, see Davachi, 2006). Moreover, based on evidence in animals and consistent with a multiple memory systems theoretical framework, it has been suggested that learning that depends on the hippocampus (but not the striatum) will result in better memory for feedback events (White and McDonald, 2002). Therefore, we tested
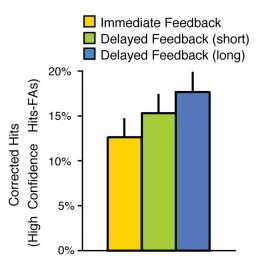

Figure 8. Episodic memory for feedback events. Later memory for feedback images was better for feedback events that were Delayed versus Immediate during the Learning phase (results from a parallel behavioral study shown here). Corrected hit rates (high confidence hits minus false alarms) are shown. Error bars represent ±1 SEM.

the prediction that delayed feedback would lead to better episodic memory for feedback events. After participants completed scanning, they were given a surprise test of their memory for the trial-unique feedback images they saw during learning. This allowed us to assess memory (later status as recognized vs forgotten) broken down by Feedback Timing (Immediate vs Delayed) during learning.

Subsequent memory for feedback images was numerically better for images that had been associated with delayed feedback during learning. However, memory for feedback was highly variable across participants and, consistent with the incidental nature of the task, was relatively low across the group. Thus, to be able to address whether delayed feedback would lead to better episodic memory for feedback images, we conducted a separate behavioral study in a larger group of participants ($n = 25$). This separate group of participants completed the exact same tasks as the scanned participants, but did so in the laboratory without undergoing fMRI scanning. As shown in Figure 8, this study confirmed the trend in the scan data and revealed that participants had significantly better memory for feedback images that had been delayed than for feedback images that were immediate (linear effect of delay: $F_{(1,24)} = 9.04$, $p = 0.006$). These results provide further evidence that different learning and memory processes are engaged when feedback is immediate versus delayed.

Discussion

Our results provide converging evidence from patient and fMRI studies in humans indicating that the striatum and hippocampus play complementary roles in learning as a function of feedback timing. Individuals with disrupted nigrostriatal function due to Parkinson's disease were impaired at learning from immediate but not delayed feedback. Using fMRI, we further found that healthy individuals showed activation in the striatum during learning from immediate feedback and in the hippocampus during learning from delayed feedback. The finding that the hippocampus supports learning from delayed feedback suggests a possible complementary mechanism for learning in the many situations in which feedback occurs with a temporal delay. Additionally, after learning, memory for delayed feedback events was better than memory for immediate feedback events, suggesting that feedback timing had consequences not just for the engage-

ment of distinct neural systems but also for the representation of what was learned. Together, these findings indicate that multiple neural systems support learning from feedback and that their contributions are modulated depending on when feedback occurs.

The striatum and immediate outcomes

The current results are consistent with extant evidence indicating that dopaminergic contributions to learning and decision making may be modulated by the timing of feedback or by the temporal framing of decisions. Electrophysiological data show that the timing of rewards modulates responses of midbrain dopamine neurons. Rewards that predictably arrive with a delay of several seconds elicit a response similar to rewards that are entirely unpredicted (Fiorillo et al., 2008; Kobayashi and Schultz, 2008), indicating that the midbrain-striatal system does not effectively learn to predict delayed rewards.

Our findings also provide a link between the role of the striatum in feedback-driven learning and in intertemporal choice. Despite debate about the precise mechanism by which the striatum contributes to intertemporal choice (Kable and Glimcher, 2009; Figner et al., 2010), converging evidence indicates that representations in the ventral striatum are modulated by whether outcomes are immediate versus delayed, and it has been shown repeatedly that the ventral striatum exhibits greater responses to the choice of sooner versus later outcomes (McClure et al., 2004; Kable and Glimcher, 2007; Roesch et al., 2007; Gregorios-Pippas et al., 2009; Luo et al., 2009; Pine et al., 2010).

The hippocampus and delayed outcomes

The finding that the hippocampus complements the striatum by supporting learning from delayed feedback contributes to a growing literature emphasizing the role of the hippocampus in binding elements across time (Cohen and Eichenbaum, 1993; Shohamy and Wagner, 2008). In humans, activation in the hippocampus is modulated by the extent to which memory encoding requires the binding of information across a gap of several seconds (Staresina and Davachi, 2009). Additionally, a recent study found that choosing delayed over immediate rewards was related to increased activation in the hippocampus (Peters and Büchel, 2010). Numerous classical conditioning experiments in animals and humans have also shown that the hippocampus is necessary when there is a temporal gap between a cue and an outcome, but not when cue and outcome are temporally contiguous (Thompson and Kim, 1996; Clark and Squire, 1998; Cheng et al., 2008).

Notably, few studies have focused on the role of the hippocampus in instrumental conditioning in animals. In general, delaying feedback in healthy animals leads to slower learning (Perin, 1943; Grice, 1948; Lattal and Gleeson, 1990; Dickinson et al., 1992; Port et al., 1993; Cheung and Cardinal, 2005). This behavioral pattern has also been demonstrated in perceptual classification learning in humans (Maddox et al., 2003; Maddox and Ing, 2005). Our findings of equivalent learning performance in immediate and delayed feedback conditions in healthy participants (replicated in three separate groups) are inconsistent with these reported findings. It is difficult to know why, but we speculate that differences in experimental design and in species may be important. For example, the studies by Maddox and colleagues (Maddox et al., 2003; Maddox and Ing, 2005) may have placed greater demands on working memory compared with our design.

Interestingly, in animals with hippocampal lesions, the behavioral impairment in learning from delayed feedback is attenuated, leading to a paradoxical effect whereby hippocampal lesions improve learning from delayed feedback (Cheung and Cardinal, 2005). Although this result appears to be in contradiction to ours, it is important to note that the "improvement" is due to the hippocampal lesions correcting for otherwise impaired learning with delayed feedback—an impairment that was not found in any of our healthy participant groups.

Together, these findings suggest that there may be basic differences in how some tasks are learned in humans versus rodents. In the rodent conditioning studies, it has been hypothesized that learning from delayed feedback is impaired because the rodents have difficulty knowing whether feedback is related to a cue, an action, or the learning environment (context) itself. Longer feedback delays exacerbate the problem. Preexposing animals to the learning context or lesioning the hippocampus, thought to be critical in encoding the context, alleviates the problem (Dickinson et al., 1992, 1996; Cheung and Cardinal, 2005). However, knowing which cue or action delayed feedback is related to is less likely to be an issue for human participants who have a better explicit understanding of the task demands.

Prediction errors in the hippocampus

Finding that the hippocampus codes for prediction errors is consistent with the proposal that the hippocampus encodes violations of expectations as shown in mnemonic contexts that do not involve reinforcement (Kumaran and Maguire, 2006, 2007; Duncan et al., 2009). By demonstrating that activation in the hippocampus varies with trial-by-trial prediction errors during learning, our findings extend the role for prediction signals in the hippocampus beyond detection and encoding of novel episodes to include feedback-driven learning of stimulus–outcome associations. Thus, the hippocampus may play a broader role in learning than previously recognized.

The detection of prediction error signals in the hippocampus, where they are not routinely reported (for a recent report, see Dickerson et al., 2011), also raises important questions about the neurobiological mechanisms underlying this signal. In the striatum, prediction error responses have been demonstrated repeatedly with fMRI and are presumed to reflect inputs from phasic firing of midbrain dopamine neurons (Schultz, 1998; D'Ardenne et al., 2008). The hippocampus is also innervated by midbrain dopamine neurons, and dopamine plays an important role in hippocampal plasticity (Otmakhova and Lisman, 1998; Shohamy and Adcock, 2010). Thus, one natural suggestion would be that the prediction error signals in the hippocampus reflect phasic dopaminergic inputs, similarly to the striatum. However, although it remains unknown precisely how dopamine modulates the hippocampus, it has recently been proposed that the hippocampus may be relatively more sensitive to tonic rather than phasic dopamine responses (for discussion, see Shohamy and Adcock, 2010). It should also be noted that the present results could not disambiguate a scalar prediction error from a signal reporting a generic mismatch between expectation and outcome. Future work is needed to fully characterize the nature of hippocampal prediction error signals and their role in learning.

Multiple brain systems for learning

The current results highlight the need for an integrated view of how brain systems contribute to learning. Most studies examining the contributions of the hippocampus have not aimed to understand its role in feedback-based learning, perhaps due to traditional framing of the roles of the striatum and hippocampus as distinct or competitive (Sherry and Schacter, 1987; Poldrack et al., 2001; Squire, 2004). Yet recent findings suggest that the hip-

pocampus may in some cases contribute to feedback-driven learning (Foerde et al., 2006; Shohamy and Wagner, 2008), outcome prediction (Johnson and Redish, 2007; Peters and Büchel, 2010), and outcome processing (Watanabe and Niki, 1985; Liu and Richmond, 2000; Wirth et al., 2009). Additionally, there has been an emerging focus on the role of dopamine in modulating episodic memory in the hippocampus (Lisman and Grace, 2005; Wittmann et al., 2005; Adcock et al., 2006; Düzel et al., 2009; Shohamy and Adcock, 2010).

Similarly, studies of feedback-driven learning have tended to focus relatively narrowly on the striatum and its dopaminergic inputs. However, recent findings suggest that feedback-based learning may involve a broader set of brain systems and cognitive processes (Doll et al., 2009; Gläscher et al., 2010). Together with the present results, these findings emphasize the need for a better understanding of how multiple learning systems interact, the contexts in which their engagement is elicited, and their relationship to behavior.

Conclusions

Our findings suggest that multiple neural systems support feedback-based learning and are modulated by feedback timing. The results further suggest that the ubiquitous finding of impaired feedback-based learning in Parkinson's disease is in fact selective to circumstances involving immediate feedback, while learning from delayed feedback is spared. In addition, our findings indicate that the ability to link feedback to an earlier action—even when there is only a short temporal gap between them—depends on computations performed in the hippocampus. Finally, the convergence of findings from patients and functional brain imaging reveal that what may appear to be qualitatively similar behavior in healthy individuals may in fact be supported by processes performed by distinct neural systems.

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.

Agid Y, Cervera P, Hirsch E, Javoy-Agid F, Lehericy S, Raisman R, Ruberg M (1989) Biochemistry of Parkinson's disease 28 years later: a critical review. Mov Disord 4 [Suppl 1]:S126–S144.

Cheng DT, Disterhoft JF, Power JM, Ellis DA, Desmond JE (2008) Neural substrates underlying human delay and trace eyeblink conditioning. Proc Natl Acad Sci U S A 105:8108–8113.

Cheung TH, Cardinal RN (2005) Hippocampal lesions facilitate instrumental learning with delayed reinforcement but induce impulsive choice in rats. BMC Neurosci 6:36.

Clark RE, Squire LR (1998) Classical conditioning and brain systems: the role of awareness. Science 280:77–81.

Cohen NJ, Eichenbaum H (1993) Memory, amnesia, and the hippocampal system. Cambridge, MA: MIT.

D'Ardenne K, McClure SM, Nystrom LE, Cohen JD (2008) BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science 319:1264–1267.

Davachi L (2006) Item, context and relational episodic encoding in humans. Curr Opin Neurobiol 16:693–700.

Daw ND (2011) Trial-by-trial data analysis using computational models. In: Affect, learning and decision making, attention and performance (pp. 3–38) (Phelps EA, Robbins TW, Delgado M, eds). New York: Oxford UP.

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

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

Dickinson A, Watt A, Griffiths WJ (1992) Free-operant acquisition with delayed reinforcement. Q J Exp Psychol B 45B:241–258.

Dickinson A, Watt A, Varga ZI (1996) Context conditioning and free-operant acquisition under delayed reinforcement. Q J Exp Psychol B 49B:97–110.

Doll BB, Jacobs WJ, Sanfey AG, Frank MJ (2009) Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. Brain Res 1299:74–94.

Duncan K, Curtis C, Davachi L (2009) Distinct memory signatures in the hippocampus: intentional states distinguish match and mismatch enhancement signals. J Neurosci 29:131–139.

Düzel E, Bunzeck N, Guitart-Masip M, Wittmann B, Schott BH, Tobler PN (2009) Functional imaging of the human dopaminergic midbrain. Trends Neurosci 32:321–328.

Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature 392:598–601.

Figner B, Knoch D, Johnson EJ, Krosch AR, Lisanby SH, Fehr E, Weber EU (2010) Lateral prefrontal cortex and self-control in intertemporal choice. Nat Neurosci 13:538–539.

Fiorillo CD, Newsome WT, Schultz W (2008) The temporal precision of reward prediction in dopamine neurons. Nat Neurosci 11:966–973.

Foerde K, Knowlton BJ, Poldrack RA (2006) Modulation of competing memory systems by distraction. Proc Natl Acad Sci U S A 103:11778–11783.

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

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.

Glover GH, Law CS (2001) Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. Magn Reson Med 46:515–522.

Gregorios-Pippas L, Tobler PN, Schultz W (2009) Short-term temporal discounting of reward value in human ventral striatum. J Neurophysiol 101:1507–1523.

Grice GR (1948) The relation of secondary reinforcement to delayed reward in visual discrimination learning. J Exp Psychol 38:1–16.

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.

Kable JW, Glimcher PW (2007) The neural correlates of subjective value during intertemporal choice. Nat Neurosci 10:1625–1633.

Kable JW, Glimcher PW (2009) The neurobiology of decision: consensus and controversy. Neuron 63:733–745.

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

Kobayashi S, Schultz W (2008) Influence of reward delays on responses of dopamine neurons. J Neurosci 28:7837–7846.

Kumaran D, Maguire EA (2006) An unexpected sequence of events: mismatch detection in the human hippocampus. PLoS Biol 4:e424.

Kumaran D, Maguire EA (2007) Match mismatch processes underlie human hippocampal responses to associative novelty. J Neurosci 27:8517–8524.

Lattal KA, Gleeson S (1990) Response acquisition with delayed reinforcement. J Exp Psychol Anim Behav Process 16:27–39.

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

Liu Z, Richmond BJ (2000) Response differences in monkey TE and perirhinal cortex: stimulus association related to reward schedules. J Neurophysiol 83:1677–1692.

Luo S, Ainslie G, Giragosian L, Monterosso JR (2009) Behavioral and neural evidence of incentive bias for immediate rewards relative to preference-matched delayed rewards. J Neurosci 29:14820–14827.

Maddox WT, Ing AD (2005) Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. J Exp Psychol Learn Mem Cogn 31:100–107.

Maddox WT, Ashby FG, Bohil CJ (2003) Delayed feedback effects on rule-based and information-integration category learning. J Exp Psychol Learn Mem Cogn 29:650–662.

McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004) Separate neural systems value immediate and delayed monetary rewards. Science 306:503–507.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304:452–454.

Otmakhova NA, Lisman JE (1998) D_1/D_5 dopamine receptors inhibit depo-

tentiation at CA1 synapses via cAMP-dependent mechanism. J Neurosci 18:1270–1279.

Perin CT (1943) The effect of delayed reinforcement upon the differentiation of bar responses in white rats. J Exp Psychol 32:95–109.

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

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

Pine A, Shiner T, Seymour B, Dolan RJ (2010) Dopamine, time, and impulsivity in humans. J Neurosci 30:8888–8896.

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.

Port R, Curtis K, Inoue C, Briggs J, Seybold K (1993) Hippocampal damage does not impair instrumental appetitive conditioning with delayed reinforcement. Brain Res Bull 30:41–44.

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

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.

Schultz W (1998) Predictive reward signal of dopamine neurons. J Neurophysiol 80:1–27.

Sherry DF, Schacter DL (1987) The evolution of multiple memory systems. Psychol Rev 94:439–454.

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

Shohamy D, Wagner AD (2008) Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. Neuron 60: 378–389.

Shohamy D, Myers CE, Grossman S, Sage J, Gluck MA, Poldrack RA (2004) Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. Brain 127:851–859.

Squire LR (2004) Memory systems of the brain: a brief history and current perspective. Neurobiol Learn Mem 82:171–177.

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

Sutton R, Barto AG (1998) Reinforcement learning. Cambridge, MA: MIT.

Thompson RF, Kim JJ (1996) Memory systems in the brain and localization of a memory. Proc Natl Acad Sci U S A 93:13438–13444.

Watanabe T, Niki H (1985) Hippocampal unit activity and delayed response in the monkey. Brain Res 325:241–254.

White NM, McDonald RJ (2002) Multiple parallel memory systems in the brain of the rat. Neurobiol Learn Mem 77:125–184.

Wirth S, Avsar E, Chiu CC, Sharma V, Smith AC, Brown E, Suzuki WA (2009) Trial outcome and associative learning signals in the monkey hippocampus. Neuron 61:930–940.

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.