Multiple Timescales of Memory in Lateral Habenula and Dopamine Neurons

Ethan S. Bromberg-Martin,[1,*] Masayuki Matsumoto,[1,2] Hiroyuki Nakahara,[3,4] and Okihide Hikosaka[1]
[1]Laboratory of Sensorimotor Research, National Eye Institute, Bethesda, MD 20892, USA
[2]Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan
[3]Laboratory for Integrated Theoretical Neuroscience, RIKEN Brain Science Institute, 2-1 Hirosawa Wako, Saitama, 351-0198, Japan
[4]Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Yokohama, Japan
*Correspondence: bromberge@mail.nih.gov

DOI 10.1016/j.neuron.2010.06.031

SUMMARY

Midbrain dopamine neurons are thought to signal predictions about future rewards based on the memory of past rewarding experience. Little is known about the source of their reward memory and the factors that control its timescale. Here we recorded from dopamine neurons, as well as one of their sources of input, the lateral habenula, while animals predicted upcoming rewards based on the past reward history. We found that lateral habenula and dopamine neurons accessed two distinct reward memories: a short-timescale memory expressed at the start of the task and a near-optimal long-timescale memory expressed when a future reward outcome was revealed. The short- and long-timescale memories were expressed in different forms of reward-oriented eye movements. Our data show that the habenula-dopamine pathway contains multiple timescales of memory and provide evidence for their role in motivated behavior.

INTRODUCTION

In order to make optimal decisions between options, the brain must predict each option's value based on the memory of the consequences it produced in the past. This process is thought to be crucially dependent on midbrain dopamine neurons (Wise, 2004). Dopamine neurons are activated by new information about the properties of upcoming rewards, firing a burst of spikes if the reward value is better than expected and pausing their activity if the reward value is worse than expected. In this manner, their activity resembles a "reward prediction error" indicating the difference between predicted and actual rewards (Schultz et al., 1997). These signals are translated into dopamine release in downstream brain structures, which controls motivation to seek rewards (Wyvell and Berridge, 2000) and enables synaptic plasticity to learn the reward value of behavioral actions and outcomes (Reynolds et al., 2001; Wise, 2004). Thus, the proper function of the dopamine system depends on its ability to make accurate predictions about future rewards.

How are dopamine neuron reward predictions constructed from past experience? It is known that during the early stages of learning dopamine predictions emerge in parallel with behavioral measures of reward expectation (Schultz et al., 1993; Hollerman and Schultz, 1998; Takikawa et al., 2004; Day et al., 2007; Pan et al., 2008). In addition, during expert performance at behavioral tasks, dopamine neuron activity is influenced by the memory of recently received rewards (Satoh et al., 2003; Nakahara et al., 2004; Bayer and Glimcher, 2005).

Yet several vital questions remain unanswered. First, what neural sources of input contribute to the dopamine neuron reward memory? Dopamine neurons receive reward-related input from many brain structures, including the amygdala (Lee et al., 2005), pedunculopontine tegmental nucleus (Pan and Hyland, 2005; Okada et al., 2009), and lateral habenula (Matsumoto and Hikosaka, 2007). The lateral habenula is a strong candidate for this role, because its neurons carry negative reward signals opposite to those in dopamine neurons and lateral habenula stimulation inhibits dopamine neurons at short latencies (Christoph et al., 1986; Ji and Shepard, 2007; Matsumoto and Hikosaka, 2007). However, it is unknown whether these input structures adjust their neural signals based on past rewarding experience in a manner resembling that of dopamine neurons.

Second, what determines the neural *timescale of memory*—the persistence of past outcomes in affecting future predictions? There is evidence that dopamine neurons are influenced by past reward outcomes in different ways at different stages of learning (Nakahara et al., 2004; Bayer and Glimcher, 2005; Pan et al., 2008). Theories of optimal prediction propose that the neural timescale of memory should be calibrated to match the reward statistics of the environment, based on the true predictive relationship between past and future rewards (Doya, 2002; Behrens et al., 2007) which may require a mixture of multiple memory timescales (Smith et al., 2006; Kording et al., 2007; Fusi et al., 2007; Wark et al., 2009). However, it remains unknown what timescales of memory are available to lateral habenula and dopamine neurons, whether they are selected in an adaptive manner sensitive to task demands, and how the selection process unfolds over time.

To investigate these questions, we analyzed the activity of lateral habenula and dopamine neurons recorded while monkeys performed a task in which the reward value of each trial was systematically related to the past reward history. This design made it possible to make a direct comparison between neural,

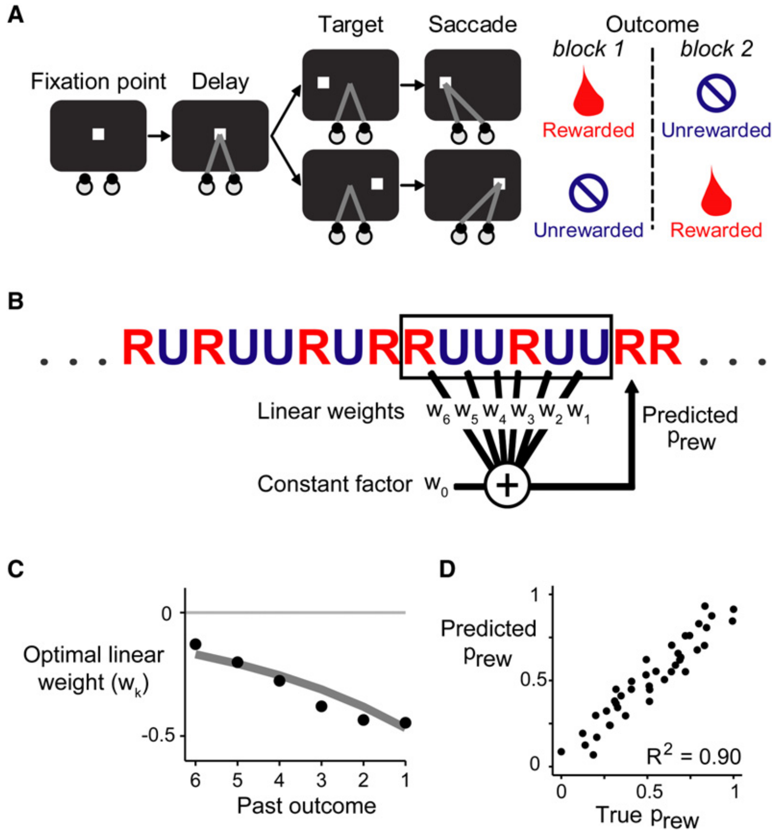

Figure 1. Behavioral Task

(A) Task diagram. The animal was required to fixate a spot of light, then follow the spot with a saccade when it stepped to the left or right side of the screen. In each block of 24 trials, saccades to one target direction were rewarded, while saccades to the other direction were unrewarded.

(B) The task used a pseudorandom reward schedule in which the reward probability could be predicted with high accuracy as a weighted linear combination of past outcomes plus a constant factor.

(C) The optimal weights (black dots) for each past reward outcome. The optimal weights were similar when constrained to take the form of an exponential decay (gray line).

(D) Plot of true reward probability against predicted reward probability using the optimal exponentially decaying linear weights. Each dot represents 1 of the 50 possible six-trial reward histories in the pseudorandom schedule. The predicted reward probability was highly correlated with the true reward probability. (See also Figure S1.)

behavioral, and task-optimal reward memories. We found that lateral habenula and dopamine neurons had similar reward memories in their phasic responses to task events, consistent with the hypothesis that the lateral habenula transmits reward memory signals to dopamine neurons. In addition, we found that lateral habenula and dopamine neurons did not use a single timescale of memory at all times during the task. Instead, they switched between two distinct memories: a suboptimal short timescale of memory expressed in response to the start of a new trial, and a nearer to optimal long timescale of memory expressed at the moment the trial's outcome was revealed. The short- and long-timescale memories were also found in specific forms of reward-oriented behavior. Our data provide evidence that the habenula-dopamine pathway can rapidly change between timescales of reward memory in a behaviorally relevant manner.

RESULTS

Behavioral Task and Optimal Timescale of Memory

We trained two monkeys to perform a reward-biased saccade task (Matsumoto and Hikosaka, 2007) (Figure 1A). Each trial began with the presentation of a fixation point at the center of a screen, where the animal was required to hold its gaze. After a 1.2 s delay, the fixation point disappeared and the animal was required to saccade to a visual target that appeared on the left or right side of the screen. Saccades to one target location were rewarded with a drop of juice. Saccades to the other target location were unrewarded but still had to be performed correctly, or else the trial was repeated. Thus, the target both instructed the location of the saccade and signaled the presence or absence of reward. The rewarded and unrewarded locations were switched after each block of 24 trials. Animals closely tracked the reward values of the targets, saccading to rewarded targets at short latencies and unrewarded targets at long latencies (Matsumoto and Hikosaka, 2007) (Figure 2B, "Target RT bias").

In this task rewarded and unrewarded trials occurred equally often, but the reward probability was not fixed at 50%; the reward probability varied from trial to trial depending on the history of previous outcomes. We used a pseudorandom reward schedule in which blocks were divided into four-trial subblocks, each containing a randomized sequence of two rewarded target trials and two unrewarded target trials. The result was that the reward sequence was more predictable than would be expected by chance: the reward probability on each trial was *inversely* related to the number of rewards that had been received in the recent past (Nakahara et al., 2004) (Supplemental Experimental Procedures). Specifically, the reward probability could be well approximated as a weighted linear combination of the previous six reward outcomes plus a constant factor (Figures 1B–1D). The optimal linear weights were largest for the most recent reward outcomes, and the weights had a negative sign reflecting the inverted relationship between past and future rewards (Figure 1C). Applying these linear weights to the true sequence of rewards in the task produced a highly accurate prediction of each trial's reward probability ($R^2 = 0.90$, Figure 1D).

The optimal linear prediction rule in this task resembles classic theories of reinforcement learning (Rescorla and Wagner, 1972; Sutton and Barto, 1981) in which past outcomes have a linear effect on future reward predictions (Sutton and Barto, 1998; Nakahara et al., 2004; Bayer and Glimcher, 2005). But there is a crucial difference. In classic theories, if a stimulus is followed

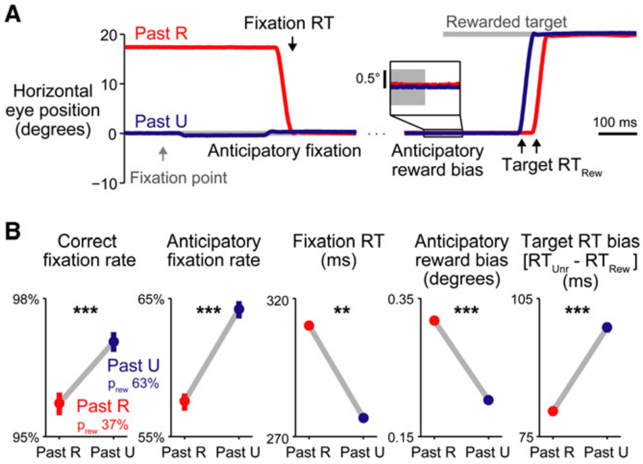

Figure 2. Behavioral Memory for a Single Previous Outcome

(A) Trace of horizontal eye position during two example rewarded trials, when the past trial was rewarded (Past R, red) or unrewarded (Past U, blue). Gray bars indicate the fixation point and saccade target. Left: eye position aligned at the time of fixation point onset. Right: eye position aligned at target onset. Inset: eye position aligned at target onset, showing a small bias in eye position toward the location of the rewarded target.

(B) Measures of behavioral performance, separately for trials when the past trial was rewarded (red) or unrewarded (blue). Target RT bias is the mean difference in reaction time between saccades to the unrewarded target versus rewarded target. Bars are 80% bootstrap confidence intervals. Asterisks indicate statistical significance. **$p < 10^{-4}$ in combined data, $p < 0.05$ in monkey L; ***$p < 10^{-4}$ in combined data, $p < 0.05$ in monkey L, $p < 0.05$ in monkey E; bootstrap test. The memory for past outcomes influenced behavioral performance at all times during the trial. (See also Figure S2.)

by reward, then this *increases* the estimated value of that stimulus in the future. Whereas in our task, if a trial is followed by reward, then this should *reduce* the estimated value of task trials in the future (for a formal model, see Figure S1). In this sense, our task may resemble a foraging situation in which collecting rewards at a foraging site reduces the number of rewards that are available at that site on future visits. We therefore set out to test whether animals and neurons could predict rewards in this ''inverted'' task environment.

Behavioral Memory for a Single Past Reward Outcome

We first analyzed the effect of a single previous reward outcome on animal behavior. The true reward probability given a single past outcome was 37% after rewarded trials and 63% after unrewarded trials. Consistent with previous studies (Nakahara et al., 2004; Takikawa et al., 2002), we found that animals used this feature of the task to predict future rewards, indicated by their improved task performance on trials when the reward probability was high (Figure 2B, ''Correct fixation rate''). In order to obtain a finer measure of how the animals' reward memory evolved over the course of each trial, we examined the time course of their eye movements. Past outcomes influenced eye movements in anticipation of each task event and in reaction to each task event (Figures 2A and S2). In anticipation of the fixation point, animals often positioned their eyes at the center of the screen in order to initiate the trial more quickly. When the reward probability was higher, they anticipated the trial more often (Figure 2B, ''Anticipatory fixation rate''). One animal was less perfect in anticipation and often had to react to the fixation point by shifting its gaze. When the reward probability was higher, its reactions to the fixation point were faster (Figure 2B, ''Fixation RT''). Then, as animals anticipated the upcoming saccade targets, their eyes drifted minutely toward the rewarded target location. This drift was stronger when the previous trial was rewarded (Figure 2B, ''Anticipatory reward bias''). Finally, when the saccade target arrived, animals reacted more quickly to the rewarded target than the unrewarded target, and when the reward probability was higher this reward-oriented saccade bias was stronger (Figure 2B, ''Target RT bias''). Thus, the animal's memory for past outcomes could be measured at the start of the trial when the fixation point appeared as well as the end of the trial when the saccade target appeared, in both anticipatory and reactive eye movements.

Neural Memory for a Single Past Reward Outcome

To examine the neural basis of the single-trial memory, we next analyzed the activity of 65 neurons recorded from the lateral habenula and 64 reward-responsive presumed dopamine neurons recorded from the substantia nigra pars compacta (Matsumoto and Hikosaka, 2007) (Experimental Procedures). Figure 3A shows the population average activity of lateral habenula neurons. These neurons carried strong negative reward signals (Matsumoto and Hikosaka, 2007). They were phasically inhibited by the cue signaling the start of a new trial (''fixation point'') and the cue signaling reward (''rewarded target'') but were excited by the cue signaling reward omission (''unrewarded target''). Figure 3B shows the population average activity of dopamine neurons. Their response pattern was a mirror-reversal of that seen in lateral habenula neurons (Matsumoto and Hikosaka, 2007): they were excited by trial-start and reward cues and inhibited by reward-omission cues.

Thus, both populations of neurons carried strong signals predicting reward outcomes in the future; how might they be influenced by the memory of outcomes received in the past? Current computational theories of dopamine activity make a strong prediction. These theories interpret dopamine neuron activations as ''reward prediction errors'' signaling changes in a situation's expected value (Montague et al., 1996; Schultz et al., 1997; Montague et al., 2004). This theoretical account is schematically illustrated in Figure 3C and explained in detail below (see Figure S1 for a formal model and Figure S3 for single neuron examples).

During the long and variable duration of the intertrial interval, the animal's reward expectation was presumably low because the animal did not know when the next trial would begin. When the fixation point appeared it signaled a new chance to get rewards, which would cause the animal's reward expectation to rise, a positive prediction error. This inhibited lateral habenula neurons and excited dopamine neurons (Figure 3, fixation point). The prediction error was more positive when the trial's reward probability was higher (Satoh et al., 2003) (Figure 3C), and accordingly habenula neurons were more inhibited and dopamine neurons were more excited.

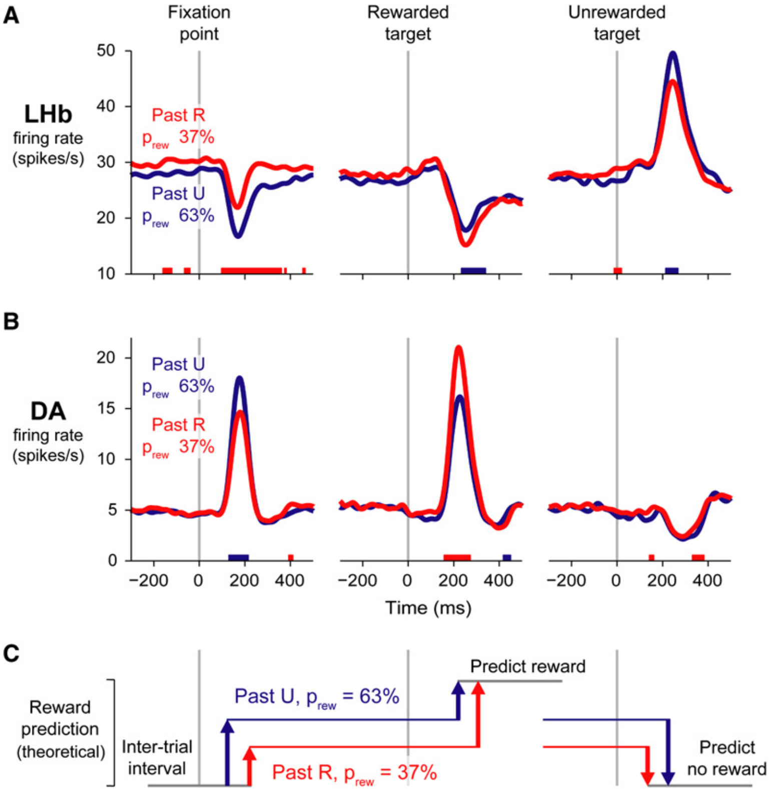

Figure 3. Neural Memory for a Single Previous Outcome

(A) Population average firing rate of lateral habenula neurons (LHb) when the past trial was rewarded (red) or unrewarded (blue). Firing rates were smoothed with a Gaussian kernel (σ = 15 ms). Colored bars on the bottom of each plot indicate times when the past trial outcome had a significant effect on neural activity ($p < 0.01$, paired Wilcoxon signed-rank test).

(B) Same as (A), for dopamine neurons (DA). Lateral habenula and dopamine neurons had opposite mean response directions and opposite past-outcome effects during all three task events.

(C) Schematic illustration of theoretical reward predictions at each time during the trial (see text for full description). When the reward prediction increased (upward arrows, positive prediction errors), lateral habenula neurons were inhibited and dopamine neurons were excited; when the reward prediction decreased (downward arrows, negative prediction errors), lateral habenula neurons were excited and dopamine neurons were inhibited. (See also Figure S3.)

If the fixation point was followed by the rewarded target, the reward expectation would rise further up to 100%, a second positive prediction error. This again inhibited lateral habenula neurons and excited dopamine neurons. In this case, however, the prediction error was *less* positive when the trial's reward probability was higher, because the high initial expectation only needed to be increased by a small amount to reach its maximal level (Figure 3C). Indeed, when the reward probability was higher, habenula neurons were less inhibited, and dopamine neurons were less excited (Figure 3, rewarded target).

Finally, if the fixation point was followed by the unrewarded target the reward expectation would fall to 0%, a negative prediction error. This excited lateral habenula neurons and inhibited dopamine neurons. The prediction error was more negative when the trial's reward probability was higher, because the high initial expectation had to fall farther to reach its minimal level (Figure 3C). Indeed, when the reward probability was higher, habenula neurons were more excited and dopamine neurons were more inhibited (Figure 3, unrewarded target). The reward probability effect was rather weak for dopamine neurons, presumably because their firing rate on unrewarded trials was close to zero and had little room to be modulated by reward expectation (Bayer and Glimcher, 2005) (Figure 3B).

In summary, lateral habenula and dopamine neurons had opposite phasic past-outcome effects to match their opposite direction of phasic responses, consistent with the hypothesis that the lateral habenula transmits reward memory signals to dopamine neurons.

Multiple Timescales of Memory

We next asked how far the neural memories extended into the past, and whether they remained consistent over the course of the trial. In particular, the theoretical "reward prediction error" model in Figure 3C implies that all neural responses during the trial should have the same timescale of memory, because the responses should be based on the same neural prediction about the trial's reward value (Figure S1). To test this, we fit the firing rates of each neural population as a linear combination of past reward outcomes (Bayer and Glimcher, 2005). To reduce the number of fitted parameters, we used a model in which all neurons in a population shared the same timescale of memory but each neuron could carry the memory signal to a greater or lesser degree (for example, due to differences in response gain). Thus, the single-trial neural firing rates were fit by the equation:

$$\text{rate}_{n,t} = \mu_n + a_n(\beta_1 r_{t-1} + \beta_2 r_{t-2} + \beta_3 r_{t-3} + \ldots + \beta_6 r_{t-6}) + N(0, \sigma_n),$$

where $\text{rate}_{n,t}$ is the firing rate of neuron n on trial t, μ_n is the neuron's mean firing rate, a_n is the neuron's "memory amplitude" (strength of memory effects), β_k is the population's "memory weight" for the outcome received k trials ago, $r_{t\text{-}k}$ is the reward outcome k trials ago (+0.5 if rewarded, -0.5 if unrewarded), and σ_n is the neuron's spiking noise (standard deviation of the firing rate).

In this model, the relative influence of each past outcome was controlled by the memory weight vector β, a parameter shared among all neurons, while the magnitude and direction of memory effects were controlled by the memory amplitudes a_n, which were specific to each neuron. Using this model, we estimated the average effect of each past outcome on the firing rate. For each past outcome k, the effect was equal to the memory weight β_k multiplied by the population average of the memory amplitudes a_n, yielding the change in firing rate caused by the outcome received k trials ago ("Past Rewarded – Past Unrewarded,"

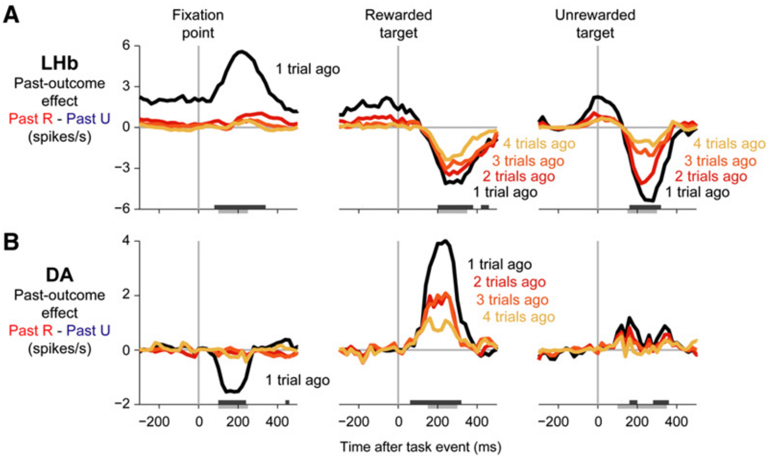

Figure 4. Multiple Timescales of Memory

(A and B) Memory effects in lateral habenula neurons (A) and dopamine neurons (B). Each panel shows the population average past-outcome effects—the difference in firing rate depending on whether a past outcome was rewarded or unrewarded ("Past R – Past U"), derived from the parameters of the fitted model described in the main text. Colored lines are the firing rate differences for specific past outcomes (black, red, orange, yellow = one, two, three, four trials-ago outcomes). The analysis was performed in a 151 ms sliding window advanced in 20 ms steps. Dark gray bars at the bottom of the plot indicate times when the population average memory amplitude was significantly different from zero, using the version of the memory model in which the weights followed an exponential decay ($p < 0.01$, Wilcoxon signed-rank test). Light gray bars below the axes are the time windows used for the analysis in Figure 5. Both lateral habenula and dopamine neurons had one-trial memories in response to the fixation point, but multiple-trial memories in response to the targets. (See also Figure S4.)

Figure 4). We then calculated the past-outcome effect at each time point during the trial by fitting the model in a sliding window advanced over the entire neural response (Figure 4).

Neurons had strikingly different timescales of memory at different times during the trial (Figures 4A and 4B). In response to the onset of the fixation point, both lateral habenula and dopamine neurons had a short timescale of memory, primarily influenced by only a single previous reward outcome. However, in response to the targets their memory suddenly *improved*, taking on a long timescale of memory with a strong influence of at least three previous outcomes. Analysis of single-neuron activity showed that both short and long timescales of memory were present in the same population of neurons (Figure S4).

To make a quantitative comparison between the neural memories, we constrained the population memory weights β to take the form of an exponential decay, so that the memory length could be described by a single parameter, the decay rate D (Figures 5A and 5B, solid lines). The decay rate D takes on values between 0 and 1 and represents the fraction of each past outcome's influence that fades away after each trial, analogous to the learning rate parameter α used in temporal-difference algorithms for reinforcement learning (Bayer and Glimcher, 2005; Sutton and Barto, 1998). Note that this parameter does not distinguish whether neural memories decayed as a function of elapsed time or of elapsed task trials. The resulting exponentially decaying memory weights were close to the original fit in which the weights were allowed to vary independently (Figures 5A and 5B, compare solid lines to filled circles; see Table S1 for all fitted decay rates).

For habenula neurons, the memory decay rate was significantly higher for the response to the fixation point than for the responses to the rewarded target (bootstrap test, $p < 10^{-4}$) and the unrewarded target ($p = 0.03$). For dopamine neurons, the decay rate was higher for the fixation point than for the rewarded target ($p = 0.006$); a similar trend was evident for the unrewarded target, but did not reach significance ($p = 0.33$) possibly due to the lower firing rates and smaller absolute memory effects on those trials. The decay rates for the rewarded and unrewarded targets were not significantly different from each other in either population (habenula $p = 0.12$, dopamine $p = 0.39$), so for further analysis the data from both targets were pooled by fitting them with a single decay rate (Experimental Procedures).

We next compared the memory timescales found in neural activity with the memory timescale of the task-optimal reward prediction rule (gray curve, Figure 1C). All neural responses had significantly higher decay rates than the optimal predictor, indicating that they all had a shorter-than-optimal timescale of memory (all $p < 0.05$; see also Figure 7). The optimal timescale was approached most closely by the long-timescale neural responses to the target, suggesting that the neural responses to the target were most closely matched to the reward statistics of the task.

To understand the functional significance of the neural timescales of memory, we compared them to the behavioral timescales of memory seen in anticipatory eye movements and saccadic reaction times (Figures 5C and 5D). These were fitted using the same procedure that was used for neural activity, producing a comparable set of memory weights (Experimental Procedures). This analysis produced two main results. First, anticipatory eye movements had a long timescale of memory at all times during the trial, both in anticipation of the fixation point and of the target (Figure 5C). Both types of anticipatory eye movements had a longer timescale of memory than the neural response to the fixation point (anticipation of fixation point versus neural response to fixation point: habenula $p = 0.025$, dopamine $p = 0.037$; anticipation of target versus neural response to fixation point: habenula $p < 10^{-4}$, dopamine $p = 0.002$). Thus, at the moment when the fixation point appeared neural activity was only influenced by a single past outcome even though behavioral anticipation was influenced by multiple past outcomes. This shows that neurons were not bound to follow the timescale of memory present in behavior. Consistent with this finding, a control analysis showed that neural memory effects were not simply caused by neural coding of behavioral output (Figure S5).

This raised the question of whether the neural timescale of memory could be linked to any motivational process that drove animal behavior. A second analysis, focused on reaction times,

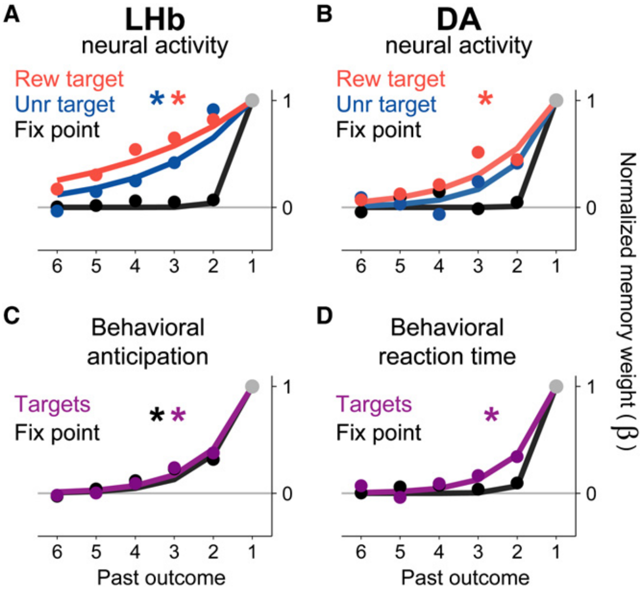

Figure 5. Quantifying Neural and Behavioral Timescales of Memory

This figure shows the fitted influence of past outcomes on the activity of lateral habenula and dopamine neurons (A and B) and on behavioral anticipatory eye movements (C) and saccadic reaction times (D).

(A) Fitted memory weights (β weights) for the lateral habenula neural population during responses to the rewarded target, unrewarded target, and fixation point (red, blue, and black). The memory weights are normalized so that $\beta_1 = 1$ (Experimental Procedures). Solid dots are memory weights from a fit in which all weights were allowed to vary independently (like those shown in Figure 4). Colored lines are a fit in which the weights were constrained to follow an exponential decay (Experimental Procedures). This analysis was done on neural activity within the time windows indicated by the gray bars below the axes in Figure 4. Asterisks indicate that the fitted memory decay rate is significantly different from 1.0 (bootstrap test, $p < 0.05$).

(B) Same as (A), but for dopamine neurons. Both lateral habenula and dopamine neurons had long-timescale memories in response to the targets, but short-timescale memories in response to the fixation point.

(C) Fitted memory weights for anticipatory behavior, separately for anticipatory fixation (black) and anticipatory bias toward the rewarded target (purple).

(D) Fitted memory weights for saccadic reaction times, separately for reactions to the fixation point (black) and targets (purple). (See also Figure S5.)

provided a possible candidate. In parallel with the pattern seen in neural activity, behavioral reaction times to the fixation point had a short timescale of memory, whereas reaction times to the targets had a longer timescale of memory (Figure 5D, $p = 0.017$). When compared to neural activity, the behavioral timescale for the fixation point was shorter than the neural timescale for the targets (habenula $p < 10^{-4}$, dopamine $p = 0.035$), and likewise, the behavioral timescale for the targets was longer than the neural timescale for the fixation point (habenula $p = 0.010$, dopamine $p = 0.028$). A caveat is that the measured timescales for reaction times were primarily dependent on one animal that had a larger amount of data (Figure S7). Taken together, these data suggest that lateral habenula and dopamine neurons do not share a common reward memory with the neural process that drives proactive, anticipatory eye movements but may share a common memory with the neural process that drives reactive, saccadic eye movements.

Timescales of Memory in Tonic Neural Activity

Our results so far suggested that the neural memory ‘‘built up’’ over time, starting each trial with a short timescale but finishing with a long timescale. If this was the case, then neural activity during the intermediate portion of each trial should have an intermediate timescale. To test this hypothesis, we checked for memory effects in tonic neural activity during the pretarget period and intertrial interval.

We found that the majority of lateral habenula neurons carried reward-related signals in their tonic activity (Figures 6A and 6C). In the example shown in Figure 6A, the neuron was phasically excited by the unrewarded target but then switched to be tonically excited after rewarded outcomes, a signal that continued during the intertrial interval and carried into the next trial (this neuron also had a second phasic excitation on unrewarded trials at the time of reward omission, a response found in a fraction of habenula neurons [Matsumoto and Hikosaka, 2007; Hong and Hikosaka, 2008] which also had a memory effect [Figure S6]).

The example habenula neuron had the most typical pattern of tonic memory effects, with tonic excitation after past rewards. However, the opposite pattern of modulation was also common. We measured each neuron’s tonic memory effects using the area under the receiver operating characteristic (ROC) (Green and Swets, 1966). The ROC area was above 0.5 if the neuron had a higher firing rate after rewarded trials, and below 0.5 if the neuron had a higher firing rate after unrewarded trials. The tonic memory effects were strong but idiosyncratic (Figure 6C) and occurred in the same neurons as phasic memory effects (Figure S6). Consistent with our hypothesis, habenula tonic activity had an intermediate timescale of memory (Figure 6D), shorter than the response to the targets (intertrial interval, $p < 10^{-4}$; pretarget period, $p = < 10^{-4}$) but tending to be longer than the response to the fixation point (intertrial interval, $p = 0.06$; pretarget period, $p = 0.009$).

Dopamine neurons could also be tonically excited or inhibited after past rewards (Figures 6B and 6E). Their past-reward effects were generally modest in size (Figure 6E) but reached significance in a much greater proportion of neurons than expected by chance (binomial test, intertrial interval $p < 10^{-12}$, pretarget period $p = 0.009$). The modest size and variable direction of these effects may explain why they have not been reported before to our knowledge. During the intertrial interval these tonic effects appeared to have a short timescale of memory, similar to the dopamine neuron response to the fixation point and shorter than in the response to the targets (Figure 6F), although the latter difference did not reach significance ($p = 0.14$). During the pretarget period their tonic effects were too weak for the timescale of memory to be estimated accurately (Table S1).

Time-Varying Changes in the Timescale of Memory

Taken as a whole, the timescales of neural memory during the task followed a V-shaped pattern (Figure 7). This was clearest in lateral habenula neurons where tonic activity was common and the ebb and flow of memory effects could be tracked during all task periods. The timescale started as a one-trial memory in response to the fixation point, lengthened during the pretarget period, reached a climax in response to the target, and then faded back to a one-trial memory again during the intertrial

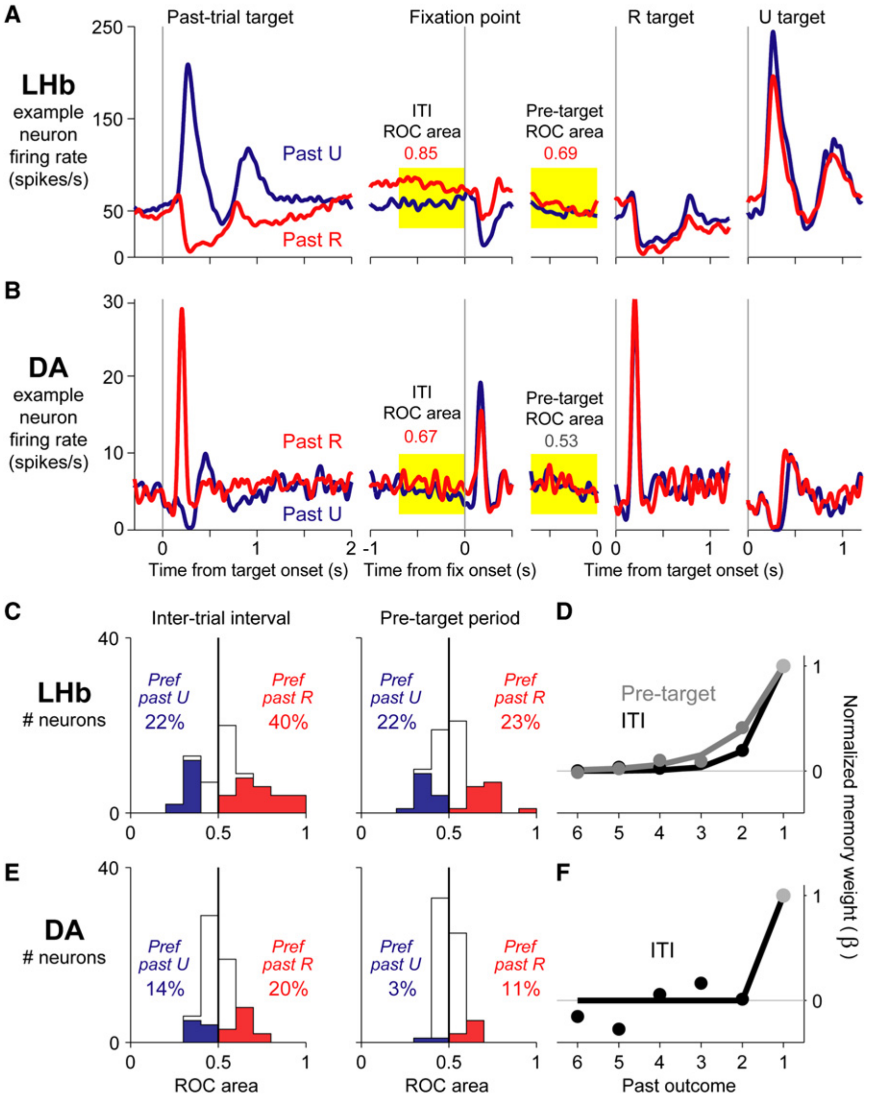

Figure 6. Timescales of Memory in Tonic Neural Activity

This figure shows the effect of a single past outcome on tonic neural activity during the intertrial interval and pretarget period, for two example neurons (A and B) and quantified for all lateral habenula and dopamine neurons (C and E). Also shown is the fitted influence of multiple past outcomes on tonic activity (D and F).

(A) Activity of an example lateral habenula neuron on rewarded (red) and unrewarded (blue) trials. The activity is shown for the response to the target (Past-trial target), and then is followed into the next trial. Tonic activity was analyzed during the intertrial interval (ITI, yellow 700 ms window before fixation point onset) and the pretarget period (Pre-target, yellow 700 ms window before target onset). Numbers indicate the neuron's ROC area for discriminating the past reward outcome. Colors indicate significance ($p < 0.05$, Wilcoxon rank-sum test).

(B) Same as (A), for a dopamine neuron.

(C) Histogram of lateral habenula neuron ROC areas for the intertrial interval and pretarget period. Numbers indicate the percentage of neurons with significantly higher activity on past-rewarded trials (red) or past-unrewarded trials (blue).

(D) Timescale of neural memory for the intertrial interval (black) and pretarget period (gray). Conventions as in Figure 5.

(E and F) same as (C and D), for dopamine neurons. Memory effects during the pretarget period were not strong enough to estimate the timescale of memory. (See also Figure S6.)

interval. The same V-shaped pattern was present in both animals (Figure S7). When considered over the course of multiple trials, this pattern implies that neural activity repeatedly changed between two different memory timescales, switching back and forth between them every few seconds.

DISCUSSION

We found that lateral habenula and dopamine neurons had mirror-reversed phasic memory effects, consistent with the hypothesis that the lateral habenula contributes to dopamine neuron reward memories. Unexpectedly, however, lateral habenula and dopamine neurons were not bound to a single reward memory but instead accessed at least two distinct memories for past rewards, a short-timescale memory expressed at the start of each trial, and a long-timescale memory expressed as the trial's reward outcome was revealed.

Functional Implications of Reward Memories

It is known that lateral habenula and dopamine neuron responses to rewarding cues and outcomes are modulated by predictions built on the basis of past experience. The neural algorithm which computes these predictions has been a topic of intense investigation (Schultz et al., 1997; Pan et al., 2005, 2008; Morris et al., 2006; Roesch et al., 2007). Conventional theories of the dopamine system suggest that reward predictions resemble an exponentially weighted average of past reward outcomes, a pattern that was seen in a previous study (Bayer and Glimcher, 2005). On the other hand, there is evidence that neural reward predictions can also be influenced by additional factors such as the number of trials since the most recent reward delivery (Satoh et al., 2003; Nakahara et al., 2004). Our task made it possible to assess the functional significance of these neural reward memories, by measuring the degree to which they are adapted to the reward statistics of the environment (via comparison with the task-optimal reward memory) and the degree to which they are linked to reward-related behavior (via comparison with the reward memories expressed in anticipatory and saccadic eye movements).

We found that the neural response to the reward-indicating target was based on a reward prediction resembling an

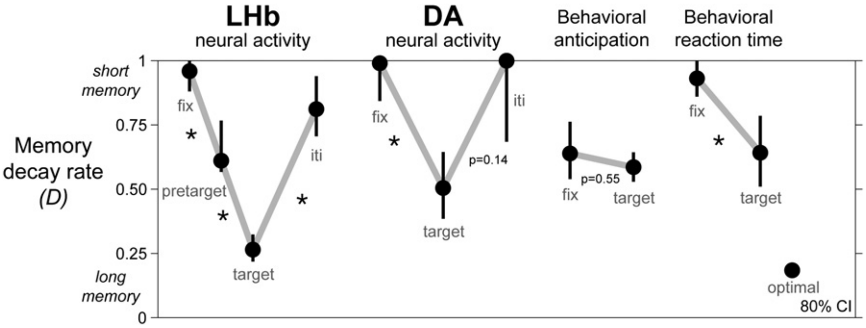

Figure 7. Time-Varying Changes in the Timescale of Memory

This figure quantifies the timescale of memory found in neural activity and behavior, separately for each lateral habenula and dopamine neuron response (LHb, DA) and for behavioral anticipatory eye movements and reaction times. Each data point for neural activity represents the fitted decay rate *D* for one of the curves shown in Figures 5A and 5B or 6D and 6F. The decay rates for behavioral anticipatory eye movements and reaction times are from Figures 5C and 5D. Far right: optimal timescale of memory (from Figure 1C). Asterisks indicate significant differences in the fitted decay rates ($p < 0.05$, bootstrap test; Experimental Procedures). Nonsignificant differences are shown as written p values. Error bars are 80% bootstrap confidence intervals. (See also Figure S7 and Table S1.)

exponentially weighted average of past outcomes, similar to the prediction rule derived from classic theories. This confirms previous findings in dopamine neurons and shows that lateral habenula neurons also signal reward predictions built by integrating multiple past outcomes. However, the neural reward predictions were related to past outcomes in a *negative* manner. This is opposite to the relationship predicted by classic theories and measured in a previous study (Bayer and Glimcher, 2005) but is similar to the rule derived for the optimal reward predictor in our task. This shows that lateral habenula and dopamine neurons integrate multiple past outcomes in a flexible manner that is tuned to the reward statistics of the task at hand.

In addition, the neural response to the target had a longer timescale of memory than the neural response to the fixation point. Indeed, the neural response to the target matched the longest timescales of memory seen in animal behavior and approached (although did not achieve) the timescale of the task-optimal prediction rule. The long timescale of memory of the target response may be a result of the target's importance for reward prediction. The target indicated the upcoming reward outcome with high accuracy, whereas the fixation point did not provide any new information about future outcomes. In other words, neurons accessed their most optimized timescale of memory at the moment when animals viewed the most informative cue for predicting future rewards. Thus, our data demonstrate a possible mechanism by which lateral habenula and dopamine neurons could respond to reward information with improved accuracy by shifting to a task-appropriate timescale of memory. Along with our own data, this mechanism may account for a puzzling observation from previous studies: that dopamine neurons encode a task trial's expected value inaccurately at the onset of the trial, but later encode its value with improved accurately when responding to new information about the trial's reward outcome (Satoh et al., 2003; Bayer and Glimcher, 2005). Given the role of dopamine in reinforcement learning (Wise, 2004), this mechanism would improve the accuracy of dopaminergic reinforcement signals at the moment when they are most needed for effective learning.

In contrast to the target response, the fixation point response had a suboptimal one-trial memory. The fixation point response did not approach the longest timescales of memory present in behavior and neural activity, and its short-timescale memory could not be predicted by current computational models of reward prediction errors (Figure S1). Instead, there was evidence that the fixation point response resembled the timescale of memory seen in saccadic reaction times at the moment the fixation point appeared. This suggests that the fixation point response may be more closely related to reward-oriented behavioral reactions than to predicted reward value. This would be sensible in our task because the fixation point caused animals to make an orienting response to initiate the trial but did not provide new information about its reward value. This is also consistent with evidence that dopamine responses in certain conditions are more closely related to orienting responses and behavioral reactions than to the expected amount of primary rewards (Ljungberg et al., 1992; Satoh et al., 2003; Matsumoto and Hikosaka, 2009a; Bromberg-Martin and Hikosaka, 2009). Notably, the nigrostriatal dopamine pathway is known to be crucial for learned orienting responses to an upcoming task trial, in a manner distinct from learned approach to reward outcomes (Han et al., 1997; Lee et al., 2005).

This distinction between the fixation point and target responses is further supported by a recent study (Bromberg-Martin et al., 2010). In that study, we found that lateral habenula and dopamine responses to a ''trial start'' cue (similar to the fixation point) were enhanced on trials when the cue triggered short-latency orienting reactions. In addition, these responses reflected motivational variables in a different manner than conventional neural responses to reward value cues. When the behavioral task was changed by replacing reward outcomes with aversive stimuli, many neurons adapted by changing their responses to reward value cues in a manner consistent with reduced reward expectation. However, animals continued to orient to the trial start cue and neurons continued to respond to the trial start cue with equal strength (Bromberg-Martin et al., 2010). Our present data complement these results by showing quantitatively that the responses to the trial start cue and reward value cues do not reflect the same expectation about the trial's reward value, and that the response to the trial start cue may be linked to the neural process that motivates orienting reactions by adapting to past outcomes with a similar timescale of memory.

Neural Mechanisms Underlying Reward Memory Signals

We found that lateral habenula neurons carried phasic reward memory signals that resembled a mirror-reversed version of the memory signals in dopamine neurons. This lateral habenula activity is likely to contribute to dopamine neuron reward memories, since lateral habenula responses to the fixation point and unrewarded target occur at shorter latencies than in dopamine neurons (Matsumoto and Hikosaka, 2007; Bromberg-Martin et al., 2010), and it is known that spikes in lateral habenula neurons induced by electrical stimulation cause dopamine neurons to be potently inhibited at short latencies (Christoph et al., 1986). However, it is also possible that reward memory signals arrive in dopamine neurons through a more complex pathway. For instance, it is possible that lateral habenula and dopamine reward memories originate from a common source, or that lateral habenula signals to dopamine neurons are modified by downstream circuitry such as inhibitory neurons in the ventral tegmental area (Ji and Shepard, 2007) and rostromedial tegmental nucleus (Jhou et al., 2009). A comprehensive test of these alternatives would require recording dopamine neuron activity while manipulating lateral habenula spike transmission through lesions or inactivation.

What is the source of the short- and long-timescale memories? One possibility is that reward memories are transmitted along a sequential pathway, from upstream brain areas → lateral habenula → dopamine neurons. Memory functions have been traditionally associated with prefrontal cortical areas where past reward outcomes are known to have a persistent influence on neural activity (Barraclough et al., 2004; Seo and Lee, 2007; Simmons and Richmond, 2008), and reward outcomes also have persistent effects in subcortical areas, including the striatum (Yamada et al., 2007). A good candidate for conveying these signals to the lateral habenula is the globus pallidus, which is known to provide the habenula with short-latency reward signals (Hong and Hikosaka, 2008). Thus, one candidate pathway for transmitting reward memory signals is prefrontal cortex → striatum → globus pallidus → lateral habenula. Another candidate is a direct projection from medial prefrontal cortex → lateral habenula, suggested by anatomical studies in rats (Greatrex and Phillipson, 1982; Thierry et al., 1983). Finally, it is also possible that lateral habenula and dopamine neurons receive reward memory signals from a common source of input to both brain regions, such as the ventral pallidum or lateral hypothalamus (Geisler and Zahm, 2005).

In order to decide between these alternatives, it will be important for future studies to record activity in multiple brain areas using the same subjects and behavioral tasks, so that the reward memories in these areas can be directly compared. Notably, one brain imaging study using punishments (aversive outcomes) found that blood-oxygen level dependent signals in the amygdala had a long timescale of memory, but during the same task signals in the fusiform gyrus had a short timescale of memory (Gläscher and Büchel, 2005). A similar approach may reveal the sources of short- and long-timescale memories in the realm of rewards. Another question for further study is whether neural memories are similar for rewards and punishments (Yamada et al., 2007). Many lateral habenula neurons and dopamine neurons respond to rewards and punishments in opposite manners as though encoding motivational value, whereas other dopamine neurons respond to rewards and punishments in similar manners as though encoding motivational salience (Matsumoto and Hikosaka, 2009a, 2009b). These distinct types of punishment-coding neurons are likely to receive input from separate neural sources, suggesting that their punishment memories may be distinct, as well.

We also found that many lateral habenula neurons and some dopamine neurons reflected past reward outcomes in their tonic activity. This is unexpected based on previous studies, which largely emphasized phasic activations to task events (but see Schultz, 1986; Fiorillo et al., 2003, 2008). These tonic signals might be sent to lateral habenula and dopamine neurons by the same brain regions that send them phasic signals in response to task events. The tonic activity might also be created within the neurons themselves as a biophysical after-effect of their phasic responses on previous trials. Regardless of its origin, an important caveat is that tonic memory effects were idiosyncratic between neurons, which would make them difficult for downstream brain areas to decode. If downstream neurons simply averaged the activity of all habenula or dopamine neurons together, then the tonic effects would largely cancel each other out, leaving only phasic signals fully intact (Figure 3).

Studies of reward history effects on neural activity have often focused on the framework of stimulus-reinforcement learning (Bayer and Glimcher, 2005; Pan et al., 2008) which can be implemented by a simple mechanism involving dopaminergic reinforcement of synaptic weights (Montague et al., 1996). By contrast, our task required animals to use a more sophisticated form of reward memory, a task-specific prediction rule based on a stored memory trace of past outcomes (Figures 1 and S1). This would allow the timescale of memory to be adapted to match the reward statistics of the task environment, perhaps including the frequency of changes and reversals in stimulus values (Behrens et al., 2007; Wark et al., 2009). It will be important to determine whether this form of memory is implemented with a similar synaptic mechanism, or whether it requires memory traces to be stored in a fundamentally different manner. Also, given that this form of memory had a potent influence on neural activity and behavior in our task, it will be important to test its influence in more conventional reward learning situations, as well.

In conclusion, we found that lateral habenula and dopamine neurons make use of multiple timescales of reward memory in a manner sensitive to task demands, expanding the set of mechanisms available to this neural pathway for guiding reward-oriented behavior.

EXPERIMENTAL PROCEDURES

General

Two rhesus monkeys, E and L, were used as subjects in this study. All animal care and experimental procedures were approved by the National Eye Institute Animal Care and Use Committee and complied with the Public Health Service Policy on the humane care and use of laboratory animals. Eye movement was monitored using a scleral search coil system with 1 ms resolution. For single-neuron recordings, we used conventional electrophysiological techniques described previously (Matsumoto and Hikosaka, 2007). All statistical tests were two-tailed unless otherwise noted.

Behavioral Task

Behavioral tasks were under the control of a QNX-based real-time experimentation data acquisition system (REX, Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health [LSR/NEI/NIH], Bethesda, MD). The animal sat in a primate chair, facing a frontoparallel screen ~30 cm from the eyes in a sound-attenuated and electrically shielded room. Stimuli generated by an active matrix liquid crystal display projector (PJ550, ViewSonic) were rear-projected on the screen. The animals were trained to perform a one-direction-rewarded version of the visually guided saccade task (Figure 1A). A trial started when a small fixation spot appeared at the center of the screen. After the animal maintained fixation in a small window around the spot for 1200 ms, the fixation spot disappeared and a peripheral target appeared at either left or right, typically 15° or 20° from the fixation spot. The animals were required to make a saccade to the target within 500 ms. Errors were signaled by a beep sound followed by a repeat of the same trial. Correct saccades were signaled by a 100 ms tone starting 200 ms after the saccade. In rewarded trials, a liquid reward was delivered which started simultaneously with the tone stimulus. The intertrial interval was randomized from 2.2 to 3.2 s or (for a small number of neurons) fixed at 2.2 s. In each block of 24 trials, saccades to one fixed direction were rewarded with 0.3 ml of apple juice while saccades to the other direction were not rewarded. The direction-reward relationship was reversed in the next block. Each block was subdivided into six four-trial subblocks, each consisting of two rewarded and two unrewarded trials presented in a random order. Transitions between blocks and between subblocks occurred with no external instruction (see Supplemental Experimental Procedures for example blocks and subblocks of trials).

Database

Our database consisted of 65 lateral habenula neurons (37 in animal L, 28 in animal E) and 64 reward-responsive presumed dopamine neurons (44 in animal L, 20 in animal E). We have previously reported other aspects of most of the behavioral sessions and neurons analyzed here (Matsumoto and Hikosaka, 2007). Lateral habenula neurons were included if they were responsive to the task. We searched for dopamine neurons in and around the substantia nigra pars compacta. Putative dopamine neurons were identified by their irregular and tonic firing around five spikes/s (range: 2.0–8.7 spikes/s), broad spike waveforms (spike duration > ~0.8 ms, measured between the peaks of the first and second negative deflections; signals bandpass-filtered from 200 Hz to 10 kHz), and response to reward-predicting stimuli with phasic excitation. Neurons that did not meet these criteria were not examined further. Recordings using similar criteria found that putative dopamine and nondopamine neurons formed separate clusters with distinct electrophysiological properties (Matsumoto and Hikosaka, 2009b).

Our analysis was limited to trials with "pure" reward histories, i.e., histories in which all trials were performed correctly and which did not include reversal trials (the first trial of a block in which the reward values of the targets were unexpectedly switched). The average number of trials meeting this criterion was 98 ± 33 for habenula neurons and 94 ± 32 for dopamine neurons (mean ± SD). There was no detectable change in memory effects related to the proximity or recency of reversal trials. The initial analysis was done using a single past reward outcome (Figures 2 and 3). The full analysis of behavioral and neural memory was done using six past-reward outcomes because beyond that point the behavioral and neural memories decayed to near zero (Figures 4–7). The results did not depend on the precise number of past outcomes that were analyzed. We observed similar behavioral results during lateral habenula and dopamine neuron recording, so their data were pooled for the behavioral analysis.

Memory Model

We fit the model of past-reward effects on neural activity using the method of maximum likelihood. For the version of the model with separate memory weights for each past trial, we used the MATLAB function "fminunc" to search for the memory weight vector β that produced the maximum likelihood fit, with $\beta_2 \ldots \beta_6$ initialized to 0.5 and β_1 held fixed at 1 so that the memory weights were automatically normalized (as shown in Figure 5). For the version of the model in which the weights were constrained to follow an exponential decay, we fit the single parameter D using a gradient descent procedure with D initialized to 0.5. The memory weight vector was determined by the equation $\beta_k = (1 - D)^{k-1}$. Fitting results did not depend on the initial settings of the parameters, and for simulated data sets the fitted value of D on average matched the true value of D (data not shown). For the plots in Figures 5 and 6, the analysis windows were chosen to include the major component of the mean neural response and of one past trial memory modulation. To pool data across rewarded and unrewarded targets (Figure 7), we allowed each neuron to have different neuron-specific parameters (μ_n,a_n,σ_n) for each target, but constrained both targets to have the same the memory weight vector β.

The confidence intervals for the D parameter (Figure 7) were calculated using a bootstrap procedure: for each population of neurons, the fitting procedure was repeated separately on 20,000 bootstrap data sets each created by resampling the neurons with replacement, creating a bootstrap distribution of fitted D values. The 80% confidence intervals were created by taking the range of the 10th to 90th percentiles of the bootstrap distribution. To compare a pair of decay rates D_1 and D_2, we calculated the difference, $D_{\mathrm{diff}} = (D_1 - D_2)$, and its bootstrap confidence interval. The decay rates were considered to be significantly different at level k if $D_{\mathrm{diff}} = 0$ was excluded by the $100 \times (1 - k)$% confidence interval.

Procedures for behavioral memories were the same as those for neural memories, except the model was used to fit behavioral measurements instead of neural firing rates (see below).

Behavioral Memory

The behavioral variables were defined as follows. The correct fixation rate was the percentage of trials in which the animal fixated the fixation point to initiate the trial and continued to fixate until the target appeared (i.e., no fixation break errors). The anticipatory fixation rate was the percentage of trials in which the animal's eye was inside the fixation window within 140 ms of fixation point onset, judged to be too fast for a reactive eye movement in these monkeys based on examination of reaction time distributions (other criteria produced similar results). The anticipatory target bias was the horizontal offset of the eye position in the direction of the rewarded target location, measured at the moment when the target appeared. The reaction time to the fixation point was the time between the onset of the fixation point and the eye entering the fixation window, excluding anticipatory fixations (RT < 140 ms, 61% of trials), and very slow fixations indicating inattention to the task rather than saccadic reactions (RT > 500 ms, <2% of trials). The reaction time to the target was the time between the onset of the target and the onset of the saccade. The reward-oriented reaction time bias was calculated from the reaction times to the rewarded and unrewarded targets, using the equation $\mathrm{RT}_{\mathrm{bias}} = (\mathrm{RT}_{\mathrm{unrewarded}} - \mathrm{RT}_{\mathrm{rewarded}})$. The behavioral analysis was based on sessions in which the relevant behavioral variable could be measured on at least 10 trials. Confidence intervals and p values were computed using a bootstrap procedure, in which the analysis was repeated on 20,000 bootstrap data sets created by resampling trials with replacement. To measure the behavioral timescale of memory (Figures 5 and 7), we used the same procedure as before except fitting behavioral measures instead of neural activity. Each behavioral session was treated as a separate "neuron," except when fitting saccadic reaction times to the targets, in which case each session was divided into four separate "neurons" representing the 2 × 2 combinations of (saccade direction) × (target reward value).

To measure the optimal timescale of memory (Figure 1D, black dots and gray line), we again used the same model, but fitted to the actual reward outcomes on each trial (+0.5 for rewarded, −0.5 for unrewarded) using a large simulated data set generated from the task's subblock-based reward schedule. This produced the optimal linear predictor of a trial's reward outcome based on the recent reward history (optimal in the sense of minimizing the mean squared error). To measure the accuracy of the optimal linear predictor, we correlated its predicted reward probability for each possible history of six past outcomes with the true reward probability for those histories (computed using a large set of simulated data). For this correlation, each history was weighed by its frequency of occurrence.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Experimental Procedures, seven figures, and one table and can be found online at doi:10.1016/j.neuron.2010.06.031.

ACKNOWLEDGMENTS

We thank K. Nakamura, B.J. Richmond, D. Lee, A. Lerchner, M. Ahmadi, P. Dayan, S. Kaveri, and D. L. Sheinberg for valuable discussions. H.N. is partially supported by JSPS KAKENHI 21300129 and MEXT KAKENHI 20020034. This research was supported by the Intramural Research Program at the National Eye Institute.

Accepted: June 8, 2010
Published: August 11, 2010

REFERENCES

Barraclough, D.J., Conroy, M.L., and Lee, D. (2004). Prefrontal cortex and decision making in a mixed-strategy game. Nat. Neurosci. *7*, 404–410.

Bayer, H.M., and Glimcher, P.W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron *47*, 129–141.

Behrens, T.E., Woolrich, M.W., Walton, M.E., and Rushworth, M.F. (2007). Learning the value of information in an uncertain world. Nat. Neurosci. *10*, 1214–1221.

Bromberg-Martin, E.S., and Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. Neuron *63*, 119–126.

Bromberg-Martin, E.S., Matsumoto, M., and Hikosaka, O. (2010). Distinct tonic and phasic anticipatory activity in lateral habenula and dopamine neurons. Neuron *67*, 144–155.

Christoph, G.R., Leonzio, R.J., and Wilcox, K.S. (1986). Stimulation of the lateral habenula inhibits dopamine-containing neurons in the substantia nigra and ventral tegmental area of the rat. J. Neurosci. *6*, 613–619.

Day, J.J., Roitman, M.F., Wightman, R.M., and Carelli, R.M. (2007). Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. Nat. Neurosci. *10*, 1020–1028.

Doya, K. (2002). Metalearning and neuromodulation. Neural Netw. *15*, 495–506.

Fiorillo, C.D., Tobler, P.N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. Science *299*, 1898–1902.

Fiorillo, C.D., Newsome, W.T., and Schultz, W. (2008). The temporal precision of reward prediction in dopamine neurons. Nat. Neurosci. *11*, 966–973.

Fusi, S., Asaad, W.F., Miller, E.K., and Wang, X.J. (2007). A neural circuit model of flexible sensorimotor mapping: learning and forgetting on multiple timescales. Neuron *54*, 319–333.

Geisler, S., and Zahm, D.S. (2005). Afferents of the ventral tegmental area in the rat-anatomical substratum for integrative functions. J. Comp. Neurol. *490*, 270–294.

Gläscher, J., and Büchel, C. (2005). Formal learning theory dissociates brain regions with different temporal integration. Neuron *47*, 295–306.

Greatrex, R.M., and Phillipson, O.T. (1982). Demonstration of synaptic input from prefrontal cortex to the habenula in the rat. Brain Res. *238*, 192–197.

Green, D.M., and Swets, J.A. (1966). Signal Detection Theory and Psychophysics (New York: Wiley).

Han, J.-S., McMahan, R.W., Holland, P., and Gallagher, M. (1997). The role of an amygdalo-nigrostriatal pathway in associative learning. J. Neurosci. *17*, 3913–3919.

Hollerman, J.R., and Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. Nat. Neurosci. *1*, 304–309.

Hong, S., and Hikosaka, O. (2008). The globus pallidus sends reward-related signals to the lateral habenula. Neuron *60*, 720–729.

Jhou, T.C., Fields, H.L., Baxter, M.G., Saper, C.B., and Holland, P.C. (2009). The rostromedial tegmental nucleus (RMTg), a GABAergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses. Neuron *61*, 786–800.

Ji, H., and Shepard, P.D. (2007). Lateral habenula stimulation inhibits rat midbrain dopamine neurons through a GABA(A) receptor-mediated mechanism. J. Neurosci. *27*, 6923–6930.

Kording, K.P., Tenenbaum, J.B., and Shadmehr, R. (2007). The dynamics of memory as a consequence of optimal adaptation to a changing body. Nat. Neurosci. *10*, 779–786.

Lee, H.J., Groshek, F., Petrovich, G.D., Cantalini, J.P., Gallagher, M., and Holland, P.C. (2005). Role of amygdalo-nigral circuitry in conditioning of a visual stimulus paired with food. J. Neurosci. *25*, 3881–3888.

Ljungberg, T., Apicella, P., and Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. J. Neurophysiol. *67*, 145–163.

Matsumoto, M., and Hikosaka, O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. Nature *447*, 1111–1115.

Matsumoto, M., and Hikosaka, O. (2009a). Representation of negative motivational value in the primate lateral habenula. Nat. Neurosci. *12*, 77–84.

Matsumoto, M., and Hikosaka, O. (2009b). Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature *459*, 837–841.

Montague, P.R., Dayan, P., and Sejnowski, T.J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci. *16*, 1936–1947.

Montague, P.R., Hyman, S.E., and Cohen, J.D. (2004). Computational roles for dopamine in behavioural control. Nature *431*, 760–767.

Morris, G., Nevet, A., Arkadir, D., Vaadia, E., and Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. Nat. Neurosci. *9*, 1057–1063.

Nakahara, H., Itoh, H., Kawagoe, R., Takikawa, Y., and Hikosaka, O. (2004). Dopamine neurons can represent context-dependent prediction error. Neuron *41*, 269–280.

Okada, K.-i., Toyama, K., Inoue, Y., Isa, T., and Kobayashi, Y. (2009). Different pedunculopontine tegmental neurons signal predicted and actual task rewards. J. Neurosci. *29*, 4858–4870.

Pan, W.X., and Hyland, B.I. (2005). Pedunculopontine tegmental nucleus controls conditioned responses of midbrain dopamine neurons in behaving rats. J. Neurosci. *25*, 4725–4732.

Pan, W.X., Schmidt, R., Wickens, J.R., and Hyland, B.I. (2005). Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. J. Neurosci. *25*, 6235–6242.

Pan, W.X., Schmidt, R., Wickens, J.R., and Hyland, B.I. (2008). Tripartite mechanism of extinction suggested by dopamine neuron activity and temporal difference model. J. Neurosci. *28*, 9619–9631.

Rescorla, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In Classical Conditioning II: Current Research and Theory, A.H. Black and W.F. Prokasy, eds. (New York: Appleton Century Crofts), pp. 64–99.

Reynolds, J.N.J., Hyland, B.I., and Wickens, J.R. (2001). A cellular mechanism of reward-related learning. Nature *413*, 67–70.

Roesch, M.R., Calu, D.J., and Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat. Neurosci. *10*, 1615–1624.

Satoh, T., Nakai, S., Sato, T., and Kimura, M. (2003). Correlated coding of motivation and outcome of decision by dopamine neurons. J. Neurosci. *23*, 9913–9923.

Schultz, W. (1986). Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. J. Neurophysiol. *56*, 1439–1461.

Schultz, W., Apicella, P., and Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. J. Neurosci. *13*, 900–913.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.

Seo, H., and Lee, D. (2007). Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. J. Neurosci. *27*, 8366–8377.

Simmons, J.M., and Richmond, B.J. (2008). Dynamic changes in representations of preceding and upcoming reward in monkey orbitofrontal cortex. Cereb. Cortex *18*, 93–103.

Smith, M.A., Ghazizadeh, A., and Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. PLoS Biol. *4*, e179.

Sutton, R.S., and Barto, A.G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. Psychol. Rev. *88*, 135–170.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement Learning: An Introduction (Cambridge, MA: MIT Press).

Takikawa, Y., Kawagoe, R., Itoh, H., Nakahara, H., and Hikosaka, O. (2002). Modulation of saccadic eye movements by predicted reward outcome. Exp. Brain Res. *142*, 284–291.

Takikawa, Y., Kawagoe, R., and Hikosaka, O. (2004). A possible role of midbrain dopamine neurons in short- and long-term adaptation of saccades to position-reward mapping. J. Neurophysiol. *92*, 2520–2529.

Thierry, A.M., Chevalier, G., Ferron, A., and Glowinski, J. (1983). Diencephalic and mesencephalic efferents of the medial prefrontal cortex in the rat: electrophysiological evidence for the existence of branched axons. Exp. Brain Res. *50*, 275–282.

Wark, B., Fairhall, A., and Rieke, F. (2009). Timescales of inference in visual adaptation. Neuron *61*, 750–761.

Wise, R.A. (2004). Dopamine, learning and motivation. Nat. Rev. Neurosci. *5*, 483–494.

Wyvell, C.L., and Berridge, K.C. (2000). Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: enhancement of reward "wanting" without enhanced "liking" or response reinforcement. J. Neurosci. *20*, 8122–8130.

Yamada, H., Matsumoto, N., and Kimura, M. (2007). History- and current instruction-based coding of forthcoming behavioral outcomes in the striatum. J. Neurophysiol. *98*, 3557–3567.