Reward Prediction Error Coding in Dorsal Striatal Neurons

Kei Oyama,¹ István Hernádi,² Toshio Iijima,¹ and Ken-Ichiro Tsutsui¹

¹Division of Systems Neuroscience, Tohoku University Graduate School of Life Sciences, Sendai 980-8577, Japan, and ²Department of Experimental Zoology and Neurobiology, University of Pécs, 7624 Pécs, Hungary

In the current theory of learning, the reward prediction error (RPE), the difference between expected and received reward, is thought to be a key factor in reward-based learning, working as a teaching signal. The activity of dopamine neurons is known to code RPE, and the release of dopamine is known to modify the strength of synaptic connectivity in the target neurons. A fundamental interest in current neuroscience concerns the origin of RPE signals in the brain. Here, we show that a group of rat striatal neurons show a clear parametric RPE coding similar to that of dopamine neurons when tested under probabilistic pavlovian conditioning. Together with the fact that striatum and dopamine neurons have strong direct and indirect fiber connections, the result suggests that the striatum plays an important role in coding RPE signal by cooperating with dopamine neurons.

Introduction

Midbrain dopamine neurons, which send many projection fibers to the striatum, have recently been found to code reward prediction error (RPE) (Schultz, 2002; Fiorillo et al., 2003; Satoh et al., 2003; Pan et al., 2005; Roesch et al., 2007), the discrepancy between the prediction and occurrence of a reward. A positive RPE typically occurs at the unexpected delivery of reward or the presentation of a reward-predicting stimulus, whereas a negative RPE typically occurs at the unexpected omission of reward. Dopamine neurons are known to show excitatory activity when the RPE is positive and inhibitory activity when the RPE is negative. The RPE signal is thought to function as a teaching signal in reward-based learning (Sutton and Barto, 1998; Reynolds and Wickens, 2002; Schultz, 2006), and in fact dopamine release can change the efficacy of synaptic connections in its target neurons (Wickens et al., 1996; Reynolds et al., 2001; Canales et al., 2002). Striatal neurons and midbrain dopamine neurons have strong anatomical connectivity through their direct reciprocal projections (Anden et al., 1964; Gerfen, 1985) as well as through indirect connections via the substantia nigra reticulata (SNr) (Hajós and Greenfield, 1994; Tepper et al., 1995; Bolam et al., 2000), and via the internal segment of the globus pallidus (GPi), lateral habenula, and rostromedial tegmental nucleus (RMTg) (Herkenham and Nauta, 1979; Bolam and Smith, 1992; Bolam et al., 2000; Jhou et al., 2009a,b). This connectivity implies that striatal neurons cooperate with dopamine neurons in coding the RPE signal. In fact, several human neuroimaging studies have demonstrated

that the striatum is activated when RPE occurs (Berns et al., 2001; McClure et al., 2003; O'Doherty et al., 2003, 2004; Haruno and Kawato, 2006). However, on the neuron level, there have recently been contradictory reports concerning the RPE coding in the striatum. Morris et al. (2004) recorded the activity of TANs (tonically active neurons), or striatal cholinergic interneurons, in monkeys and concluded that TANs in monkeys do not code RPE, whereas Apicella et al. (2009) have reported that they responded to RPE in a fashion inverse to that of dopamine neurons, that is, they show inhibitory activity for a positive RPE. Roesch et al. (2009) recorded the activity of rat striatal neurons that were apparently medium spiny projection neurons and reported that only few of them appeared to code RPE signals, whereas Kim et al. (2009) found both inverse and noninverse RPE coding in several medium spiny neurons and interneurons. The purpose of this study was to further examine whether RPE signals are coded in the striatum by using the behavioral paradigm of probabilistic pavlovian conditioning with which Fiorillo et al. (2003) examined in detail the quantitative coding of RPE signals in monkey dopamine neurons. We recorded single-unit activity of striatal neurons and midbrain dopamine neurons in head-fixed rats that had been conditioned in a probabilistic pavlovian conditioning task using auditory tone stimuli as the conditioned stimuli (CSs).

Materials and Methods

Subjects. Fourteen male albino Wistar rats weighing 220–270 g were used as subjects. They were individually housed under a 12 h light/dark cycle with light onset at 8:00 P.M. Throughout the experiments, they were treated in accordance with the National Institutes of Health *Guide for the Care and Use of Laboratory Animals* and the Tohoku University *Guidelines for Animal Care and Use*.

Apparatus. Experiments were conducted in a dimly lit sound-attenuated room. Auditory stimuli were generated by a personal computer and presented diagonally from two loudspeakers (ASP-701; Elecom) 30 cm from the head of the rat. An infrared sensor system was used to detect the spout-licking movement (for details of the apparatus settings, see Fig. 1).

Behavioral task. Before behavioral training, a head fixation device consisting of two metal tubes and a stainless-steel screw as grounding refer-

Received March 31, 2010; revised July 1, 2010; accepted July 8, 2010.

This study was supported by Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) Grants-in-Aid for Scientific Research 17680027 and 19673002 and by MEXT Grants-in-Aid for Scientific Research in Priority Areas—System Study on Higher-Order Brain Functions 17022009, 18020005, and 20019005. K.O. was supported by MEXT Global Common Operating Environment Program “Basic and Translational Research Center for Global Brain Science” at Tohoku University.

Correspondence should be addressed to Dr. Ken-Ichiro Tsutsui, Division of Systems Neuroscience, Tohoku University Graduate School of Life Sciences, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan. E-mail: tsutsui@mt.hokku.ac.jp.

DOI:10.1523/JNEUROSCI.1719-10.2010

Copyright © 2010 the authors 0270-6474/10/3011447-11\$15.00/0

ence electrode was implanted under ketamine (80.0 mg/kg) and xylazine (0.8 mg/kg) anesthesia and was fixed to the skull with dental cement. After recovery from the surgery, each rat was habituated to a half-cylinder restraining device (diameter, 8.5 cm; length, 15 cm) made of acrylic. During the task training and single-unit recordings, the rats were placed in the restraining device with the head fixed firmly and painlessly in the stereotaxic device (SR-5R; Narishige). The rats were trained with a probabilistic classical conditioning procedure. Five different auditory stimuli with the same intensity but with different frequencies ranging from 1.2 to 14 kHz (1.2, 2, 5, 9, and 14 kHz) were used as CSs indicating reward probabilities ranging from $p = 0$ to 1 ($p = 0, 0.25, 0.5, 0.75$, and 1.0). Combinations of tone frequencies and reward probabilities were varied between rats. To dissociate the neuronal activity dependent on the reward probability from that of the auditory sensory response, the combinations of tone frequency and reward probabilities were organized so that a reward probability-dependent tuning of response amplitude would appear as multiphasic tuning when plotted against log-aligned tone frequency. [The purpose of such a setting was to dissociate the reward probability-dependent activity from a pure sensory response to the auditory tone stimuli that would be a monophasic tuning (Doron et al., 2002).] In each trial, a 1.5 s CS was followed by a 0.5 s delay. Whether to give a reward immediately after this delay was determined probabilistically according to the CS. In a rewarded trial, a solenoid valve opened for 250 ms delivering 50 μ l of a sucrose solution through a spout in front of the rat's mouth. The intertrial interval (ITI) was normally set to one of six durations, each consisting of a fixed 4 s plus an exponentially distributed interval with a mean of 5 s. The exception was when an unpredicted reward was given during the ITI. In that case, the time between the end of the previous trial and the unpredicted reward and the time between the unpredicted reward and the start of the next trial were both set to one of the above regular ITI durations. Trial sequence was predetermined by a computer so that each of the five CSs and the unpredicted reward appeared twice each in a block of 10 trials. A daily session consisted of 600 trials.

Single-unit recording. The recording session began after the emergence of discriminative anticipatory licking responses during the CS and delay period. After the rats had been trained with the pavlovian conditioning task, chronic access to the brain was provided by using a second surgical procedure to open a hole on the skull and attach a recording chamber over it. The position of the hole on the skull [anteroposterior (AP) = +1.5 to −1.0 from bregma; lateral (L) = 1.5 to 4.5 from midline for striatum; AP = −4.5 to −6.0 from bregma; L = 0.5 to 3.0 from midline for substantia nigra compacta (SNc) and ventral tegmental area (VTA)] was determined according to the standard stereotaxic atlas (Paxinos and Watson, 2005). After recovery from the surgery, extracellular single-unit recordings were performed using tungsten microelectrodes with a platinized tip (1–3 M Ω measured at 1 kHz, 0.125 mm-diameter shaft; FHC) during the performance of the pavlovian conditioning task. The electrode was attached to a hydraulic microdrive (MO-15; Narishige) so that it could be advanced into the brain. Electrophysiological signals were amplified (10,000 times) and bandpass filtered (low cut, 100 Hz; high cut, 10,000 Hz) with a standard biophysical amplifier (Bio Amp A2-v6; Super Tech) and were displayed on an oscilloscope (CS-4125A; Kenwood). The amplified electrophysiological signals were also audibilized and presented to the experimenter through a head speakerphone. The action potentials of isolated neurons were sorted by a window-discriminator (DDIS-1; Bak Electronics) and were displayed on a digital storage oscilloscope (DCS-7040; Kenwood). The recorded electrophysiological signals were digitized at 25 kHz by using an analog-digital conversion interface (Power 1401; CED) and were then stored on a hard disk of a personal computer (X100; IBM). The times of the detected action potentials, licking movement, and the task events were also stored on that disk. Rasters and histograms showing the neuronal activity recorded under each probability condition and the response to the unpredicted reward were displayed on-line on a LCD video screen. If the neuronal activity was judged (from the experimenter's visual inspection) to be related to the task events (CS and/or reward), or to the unpredicted reward given during the ITI, we tried to maintain the recording for at least 15 trials per

Figure 1. Outline of the task. **A**, The behavioral apparatus. The rats were involved in a probabilistic pavlovian conditioning task with the head stabilized with a head fixation device and with body movement restricted by an acrylic half-cylinder. The auditory stimuli used in this study were generated by a personal computer and presented from two loudspeakers 30 cm from the head of the rat. A sucrose solution was given through a spout in front of the rat's mouth, and an infrared sensor system was used to detect spout-licking movements. **B**, Time sequence of task events in a trial. Lt., Left; Rt., right.

CS condition and stored the recorded data on the computer for the off-line data analysis.

Recording the activity of dopamine neurons. The recording procedure for dopamine neurons was as follows. Guided by the standard rat brain atlas (Paxinos and Watson, 2005), we stereotactically advanced the electrode into the brain region that was thought to be the VTA or SNc. Then we searched for the background multineuron activity responsive to CSs and rewards during the performance of the probabilistic pavlovian conditioning task. After finding such background activity, we started to isolate single-unit activity. A neuron was judged to be a dopamine neuron when it showed properties that correspond to those of typical dopamine neurons, such as a low basal firing rate (0.1–8.0 Hz) and an action potential with a long duration (>1.5 ms).

Analysis of neuronal activity. In this study, we defined a “phasic” response as one that shows a peak latency within 500 ms of the stimulus onset and an offset latency within 800 ms of the stimulus onset. To analyze the phasic task-related activity, we compared the activity within 500 ms after a CS or reward onset with the baseline activity (the activity within 500 ms before stimulus onset) independently for each CS condition by using Student's *t* test (paired, $p < 0.05$). The response to an unpredicted reward was tested by comparing the activity within 500 ms after the onset of the unpredicted reward to the activity within 500 ms before its onset. To analyze activity when the reward was omitted, we used a time window of within 2000 ms after the timing of the reward omission, as the inhibitory activity appeared to last for ~2000 ms in population activity of dopamine neurons. We used the moving-window method to calculate the latency of responses. A 40 ms test window was moved in 10 ms steps starting at the onset of the CS. The time of onset was defined as the center of the time window in the first of four consecutive steps showing a significant difference ($p < 0.05$, Student's *t* test) from the baseline activity within 500 ms before the CS onset. The time of offset was defined by the center of the time window in the first of four consecutive steps in which activity returned to baseline levels. The peak of response was determined as the maximum bin height in the 10-ms-bin histogram smoothed by the moving average of five bins. To test the dependency of CS and reward responses on reward probability, we first

Figure 2. Spout-licking behavior. **A**, Representative probability-dependent licking behavior emerged after extensive training. Rasters and histograms (bin width, 100 ms) show the timing of the licking movements observed in the rewarded and unrewarded trials under different reward probability conditions. Rasters and histograms are aligned to the CS onset. The horizontal bars beneath the histograms indicate the times of the CS presentation and the reward delivery. **B**, Representative licking behavior that is almost limited to the postreward epoch observed in the second half of a daily session. **C**, Average peak-normalized summed licking responses during the CS and delay period of six rats maintaining reward probability dependence in the first (closed circles) and second (open circles) halves of the daily sessions. **D**, Average peak-normalized summed licking responses during the CS and delay period of eight rats showing reward probability dependence in the first half of the daily sessions (closed circles) but not in the second half (open circles). Error bars show SEM.

compared the response magnitudes recorded under two extreme conditions: 100 and 0% for a CS response, and 100% and unpredicted reward for a reward response ($p < 0.05$, Student's t test). For the reward response, an unpredicted reward was regarded as a 0% condition. Then we used linear regression analysis to examine whether the responses were positively or negatively correlated with reward probability ($p < 0.05$).

Histology. Electrolytic lesions were made by passing electrical currents through the tip of the electrode and into the brain tissue. After the rat was killed with an overdose of pentobarbital, it was perfused first with 0.9% saline and then with 10% formalin before its brain was removed from the skull and stored in a 10% formalin solution. For histological inspection, the brain was sliced into 50 μ m coronal sections and stained with thionine. Slices were examined under a microscope to verify lesion site and electrode tracks. Electrode placements were classified using the rat brain atlas (Paxinos and Watson, 2005). The plots of recording sites were superimposed at 0.5 mm intervals on the coronal section of the left hemisphere.

Results

We trained 14 rats in a probabilistic pavlovian conditioning task with their head stabilized with a head fixation device and their body movement restricted by an acrylic half-cylinder (Fig. 1A). Five different auditory stimuli were used as CSs, indicating reward probabilities of 0, 25, 50, 75, and 100%. In each trial, the CS lasted 1.5 s and was followed by a 0.5 s delay, after which in the rewarded trials a sucrose solution was delivered from a spout in front of the rat's mouth (Fig. 1B). In the unrewarded trials, no sucrose solution was given at the end of the delay period. Occa-

sionally, an unpredicted reward (one not signaled by a preceding CS) was given out of the task context during the ITI.

Spout-licking behavior

We monitored spout-licking behavior during the task performance by using an infrared sensor. The licking movements were detected as interruptions of the infrared beam and were monitored throughout the training and the recording sessions. After extensive training, rats exhibited probability-dependent conditioned spout-licking movements (i.e., longer summed duration of licking for higher reward probability) during the CS and/or delay period. Figure 2A shows the licking movements that one rat exhibited under each reward probability condition. Like this one, all 14 rats showed probability-dependent licking movements during the CS and/or delay period in the first half of the daily session. The overall amount of licking in a trial gradually decreased as trials continued, perhaps because of a gradual lowering of motivational level. Figure 2B shows a typical example of sparse and probability-nondependent licking movements during the CS and delay period. For six of the rats, the licking during the second half of the daily session was less frequent than that during the first half but was still probability dependent (Fig. 2C), but for the other eight rats the infrequent licking during the second half of the daily session was completely nondependent on the reward

probability (Fig. 2D). The timing of the licking movements during the CS and delay period varied between rats (see supplemental Fig. S1, available at www.jneurosci.org as supplemental material). The continual licking movements after the reward delivery were probability-nondependent and consistent between reward probability conditions throughout the daily session in all rats.

RPE coding in striatal neurons

After the completion of training, we performed single-unit recording in the striatum during the performance of this task. Of the 984 striatal neurons whose activity was isolated, 9% ($N = 87$) showed significant phasic excitatory responses to unpredicted rewards given during ITIs ($p < 0.05$, Student's t test). (In this study, we defined a "phasic" response as one that shows a peak latency within 500 ms of the stimulus onset and an offset latency within 800 ms of the stimulus onset. See Materials and Methods for the details.) Then we tested for positive RPE coding in these neurons by comparing the responses to unpredicted rewards with the responses to rewards preceded by a CS that indicated 100% reward probability. Of the 87 neurons showing significant phasic responses to unpredicted rewards given during ITIs, 54% (47 of 87) showed significantly greater responses ($p < 0.05$, Student's t test) to rewards given during ITIs than to rewards preceded by a CS indicating 100% reward probability and were classified as "RPE-coding striatal neurons."

Figure 3A shows the activity of a representative RPE-coding striatal neuron, and Figure 3B shows the average trace of the recorded action potentials of this neuron. This neuron showed significantly stronger responses ($p < 0.0001$, Student's t test) to rewards given during ITIs than to rewards preceded by a CS indicating 100% reward probability. This neuron also showed significant phasic responses to the CSs, and the response to a CS indicating 100% reward probability was stronger than the response to a CS indicating that no reward would be forthcoming ($p < 0.0001$, Student's t test). Thus, this neuron was found to code positive RPE at both the CS and the reward onset. Linear regression analysis revealed that, for this neuron, the CS response was positively correlated with the reward probability ($r = 0.59$; $p < 0.0001$) (Fig. 3C) and the reward response was negatively correlated with the reward probability ($r = -0.74$; $p < 0.0001$) (Fig. 3D), indicating that this neuron codes positive RPE in a parametric manner.

Similar to this representative neuron, 87% (41 of 47) of all RPE-coding striatal neurons showed significant phasic responses to the CSs ($p < 0.05$, Student's t test), and 80% (33 of 41) of those neurons showed a greater CS response when the CS indicated 100% reward probability than when it indicated that no reward would be forthcoming. Of the 33 RPE-coding striatal neurons coding RPE at the CS onset, 97% ($N = 32$) showed CS responses positively correlated ($p < 0.05$) with the reward probability, and the reward responses of all 47 of the RPE-coding striatal neurons were negatively correlated ($p < 0.05$) with the reward probability.

In addition to testing the striatal neurons for positive RPE coding, we tested them for negative RPE coding. The activity at reward omission (during the first 2 s after the end of the delay period) was compared between the 75 and 0% reward probability conditions. In the 75% condition, negative RPE would occur at the reward omission because the level of reward expectation would be high after the CS onset, but in the 0% condition no negative RPE would occur at the reward omission because there would be no expectation of a reward after the CS onset. We would therefore expect negative RPE coding to be indicated by lower activity at the reward omission under the 75% reward probability condition than under the 0% reward probability condition. Only 2 of 984 striatal neurons (0.2%) showed significantly lower activity (stronger inhibition of activity) at reward omission under the 75% reward probability condition than under the 0% reward probability condition ($p < 0.05$, Student's t test). (One of these two neurons had been also

Figure 3. Activity of a representative RPE-coding striatal neuron. **A**, Rasters and histograms showing the activity recorded under different reward probability conditions. Rasters and histograms are aligned to the CS onset. The horizontal bars beneath the histograms indicate the times of the CS presentation and the reward delivery. **B**, Average trace of the recorded action potentials ($N = 100$) of this representative neuron. **C**, Discharge rate of CS responses within 500 ms after CS onset averaged over trials and plotted against the reward probability. The CS response shows a positive correlation with the reward probability ($p < 0.0001$; $r = 0.59$). **D**, Discharge rate at reward delivery (filled circles) and at reward omission (open circles) averaged over trials and plotted against the reward probability. The time window for the response to reward delivery was the first 500 ms after the reward delivery, and that for the response to reward omission was the first 2000 ms after the end of the delay period. The reward response shows a negative correlation with the reward probability ($p < 0.0001$; $r = -0.74$). In the 0% reward probability condition, the response to an unpredicted reward was used as a substitute for the reward response. Error bars show SEM.

judged to code positive RPE with its phasic responses to CSs and rewards.) For these two neurons, the activity level at reward omission was negatively correlated, or the strength of inhibitory response was positively correlated, to the reward probability ($p < 0.05$), indicating parametric coding of negative RPE.

As it is known that striatum is involved in motor functions and many of its neurons discharge in relation to movement (Rolls

Figure 4. Recording sites superimposed on the standard brain atlas. **A**, Numbers of neurons isolated and recorded within each $1 \times 1 \times 0.5$ mm tissue block within the striatum. Areas in different densities show the number of the recorded neurons. **B**, Plot of recording sites of RPE-coding striatal neuron. Numbers below each section indicate the anteroposterior coordinates (in millimeters) from bregma. The filled circles indicate the sites at which the activity of RPE-coding striatal neurons with phasic CS response was recorded, and the open circles indicate the sites at which the activity of neurons without phasic CS response was recorded.

et al., 1983; Hikosaka et al., 1989; Alexander and Crutcher, 1990), it is important to know whether the observed RPE coding is movement related. The activity of 21 of all 33 striatal neurons coding RPE at the CS onset was recorded when there was no reward probability-dependent licking during the CS and delay period (for a typical example of the absence of the reward probability-dependent licking during the CS and delay period, see Fig. 2B). It is obvious that the probability-dependent activity of these neurons is unrelated to licking movements. The activity of the other 12 striatal neurons coding RPE at the CS onset was recorded when there was probability-dependent licking during the CS and delay period (for a typical example of the reward probability-dependent licking movements during the CS and delay period, see Fig. 2A), but we confirmed that these CS responses and licking movements are not directly related as follows. On a trial-by-trial basis, we tested the correlation of the firing rate within 500 ms after the CS onset with the summed duration of licking within 2 s after the CS onset. For only two of those neurons did we find a correlation between the CS responses and the licking movements. Similarly testing the correlation of the phasic reward response with the licking movements, we found that for only 3 of 47 RPE-coding striatal neurons was the discharge rate within 500 ms after reward onset correlated with the summed duration of licking within 2 s after the reward onset. Thus, we found little relationship between the licking movements and the phasic CS and reward responses of RPE-coding neurons.

As it is known that striatum receives various sensory inputs, it is also important to confirm that the observed RPE coding cannot be explained by the frequency tuning typical of neurons in

auditory-related areas (Doron et al., 2002). In designing the task, we determined the combination of the tone frequency and the reward probability so that the reward probability-dependent response would not appear as the monophasic tone tuning. In supplemental Figure S2 (available at www.jneurosci.org as supplemental material), the magnitude of the CS response of the same RPE-coding striatal neuron shown in Figure 3 is plotted against the logarithmically scaled auditory tone frequency. To test the monophasic tone tuning, we conducted Gaussian curve fitting to the response magnitude, which resulted in a poor fitting (goodness-of-fit $p > 0.1$). We performed this test for the phasic CS responses of RPE-coding striatal neurons ($N = 41$), and all of them showed poor fitting of a Gaussian curve to the response magnitude plotted against the logarithmic scaled tone frequency. Thus, we think there is little possibility that the recorded reward probability-dependent CS responses of the RPE-coding striatal neurons are an artifact of the simple auditory tone tuning.

In the probabilistic pavlovian conditioning, uncertainty for reward, or variance of outcome, is highest in the 50% reward probability condition, gradually decreases as the reward probability increases or decreases, and reaches zero in the 0 or 100% condition. To examine

whether any of the recorded striatal neurons coded uncertainty information, we tested whether their activity was higher in the 50% reward probability condition than in the 0 and 100% condition (Student's t test, $p < 0.05$, corrected) within 500 ms after the CS onset and between 500 ms and 2 s after the CS onset. In neither time period was a neuron found to show such selectivity. Thus, we did not find any trace of uncertainty coding in the recorded striatal neurons.

The recording sites of striatal neurons were reconstructed from histological work and superimposed onto the sections of the left hemisphere of the standard rat brain atlas (Paxinos and Watson, 2005). Figure 4A shows the numbers of neurons isolated and recorded within each $1 \times 1 \times 0.5$ mm tissue block within the striatum. Most of the neurons whose activity was recorded were in the anterior and posterior parts ($AP = +1.5$ to -1.0 from bregma) of the dorsal striatum. Figure 4B shows the recording site of each RPE-coding neuron. The RPE-coding neurons were widely distributed within the dorsal striatum, and no specific topographical distribution within it was evident.

RPE coding in dopamine neurons

To directly compare the activity of RPE-coding striatal neurons with that of dopamine neurons, we recorded the activity of mid-brain dopamine neurons in the same animals in which we recorded activity in the striatum. We recorded the activity of 51 putative dopamine neurons in the SNc and the VTA (for the details concerning recording of activity of dopamine neurons, see Materials and Methods and supplemental Fig. S5, available at www.jneurosci.org as supplemental material). Of these, 59% (30

of 51) showed significant phasic responses to unpredicted rewards ($p < 0.05$, Student's t test), and 90% (27 of 30) of them showed significantly greater responses to unpredicted rewards than to fully predicted rewards ($p < 0.05$, Student's t test), indicating positive RPE coding.

Figure 5A shows the activity of a representative RPE-coding dopamine neuron, and Figure 5B shows the average trace of the recorded action potentials of this neuron. This neuron showed significantly stronger responses ($p < 0.0001$, Student's t test) to rewards given during ITIs than to rewards preceded by a CS indicating 100% reward probability. This neuron also showed significant phasic responses to the CSs, and the response to a CS indicating 100% reward probability was stronger than the response to a CS indicating that no reward would be forthcoming ($p < 0.0001$, Student's t test). Thus, this neuron was found to code positive RPE at both the CS and the reward onset. The CS response of this neuron was positively correlated with the reward probability ($r = 0.70$; $p < 0.0001$) (Fig. 5C), whereas the reward response of this neuron was negatively correlated with the reward probability ($r = -0.89$; $p < 0.0001$) (Fig. 5D, filled circles), indicating parametric coding of positive RPE. In addition, the activity level after the time of the reward omission was higher in the 0% condition than in the 75% condition ($p < 0.05$, Student's t test) and was negatively correlated with the reward probability ($r = -0.36$; $p < 0.05$) (Fig. 5D, open circles), indicating parametric coding of negative RPE.

Like this neuron, all other RPE-coding dopamine neurons showed significant ($p < 0.05$, Student's t test) phasic responses to the CSs. And in 85% (23 of 27) of these, the CS response was greater when the CS indicated 100% reward probability than when it predicted no reward. The CS responses of all 23 RPE-coding dopamine neurons coding RPE at the CS onset were positively correlated ($p < 0.05$) with the reward probability, and the reward responses of all 27 RPE-coding dopamine neurons coding RPE at the reward delivery were negatively correlated ($p < 0.05$) with the reward probability, indicating parametric coding of positive RPE. At the reward omission, 12% (6 of 51) showed significantly lower activity (stronger inhibition of activity) under the 75% reward probability condition than under the 0% reward probability condition ($p < 0.05$, Student's t test), indicating negative RPE coding. All six of these neurons had also been judged from their phasic responses to rewards to code positive RPE. For five of these neurons (83%), the activity level was negatively correlated, or the strength of inhibitory response was positively correlated, to the reward

Figure 5. Activity of a representative RPE-coding dopamine neuron. **A**, Rasters and histograms showing the activity recorded under different reward probability conditions. Alignment and notations are the same as in Figure 3. **B**, Average trace of the recorded action potentials ($N = 100$) of this representative neuron. **C**, Discharge rate of CS responses within 500 ms after CS onset averaged over trials and plotted against the reward probability. The CS response shows a positive correlation with the reward probability ($p < 0.0001$; $r = 0.70$). **D**, Discharge rate at the reward delivery (filled circles) and omission (open circles) averaged over trials and plotted against the reward probability. The time window for the response to reward delivery was the first 500 ms after the reward delivery, and that for the response to reward omission was the first 2000 ms after the end of the delay period. The reward response shows a negative correlation with the reward probability ($p < 0.0001$; $r = -0.89$). In the 0% reward probability condition, the response to an unpredicted reward was used as a substitute for the reward response. The activity after the time of the reward omission shows a negative correlation with the reward probability ($p < 0.05$; $r = -0.36$). Error bars show SEM.

probability at the reward omission ($p < 0.05$), indicating parametric coding of negative RPE.

To examine whether any of the recorded dopamine neurons coded uncertainty information, we tested whether they were more active in the 50% reward probability condition than in the 0% and 100% conditions (Student's t test, $p < 0.05$, corrected) within 500 ms after the CS onset and between 500 ms and 2 s after the CS onset. In neither time period was a neuron found to show

Figure 6. Baseline-subtracted and peak-normalized population activity of RPE-coding striatal neurons. **A**, Population histograms of RPE-coding striatal neurons ($N = 47$) around the CS presentation (left) and the timing of the reward delivery (top right) or omission (bottom right). The lines of different colors show the neuronal activity recorded under different reward probabilities (red, 100%; orange, 75%; purple, 50%; green, 25%; blue, 0%) and that recorded at the delivery of unpredicted rewards (light blue). **B**, Baseline-subtracted and peak-normalized CS responses within 500 ms after CS onset plotted against the reward probability. The CS response shows a positive correlation with the reward probability ($p < 0.0001$; $r = 0.43$). **C**, Baseline-subtracted and peak-normalized reward responses at the reward delivery (filled circles) and omission (open circles) plotted against the reward probability. The time window for the response to reward delivery was the first 500 ms after the reward delivery, and that for the response to reward omission was the first 2000 ms after the reward omission. The reward response shows a negative correlation with the reward probability ($p < 0.0001$; $r = -0.49$). In the 0% reward probability condition, the response to an unpredicted reward was used as a substitute for the reward response. Error bars show SEM.

such selectivity. Thus, we did not find any trace of uncertainty coding in the recorded dopamine neurons.

Comparison of the activities of RPE-coding striatal and dopamine neurons

To compare the activity of RPE-coding striatal neurons and that of RPE-coding dopamine neurons, we produced population average histograms of their activity (Figs. 6, 7). Figure 6A shows population average histograms of the 47 RPE-coding striatal neurons. The CS response was positively correlated with the reward probability ($r = 0.43$; $p < 0.0001$) (Fig. 6B) and the reward response was negatively correlated with the reward probability ($r = -0.49$; $p < 0.0001$) (Fig. 6C). Thus, we confirmed in the population activity the high-precision parametric coding of positive RPE in striatal neurons. No inhibitory response to reward omission (negative RPE coding) was detected in this population average activity. Figure 7A shows population average histograms of the 27 RPE-coding dopamine neurons. The CS response was

positively correlated with the reward probability ($r = 0.60$; $p < 0.0001$) (Fig. 7B) and the reward response was negatively correlated with the reward probability ($r = -0.59$; $p < 0.0001$) (Fig. 7C, filled circles). In this population-level data, we could also confirm that the activity after the time of the reward omission was negatively correlated with the reward probability ($r = -0.44$; $p < 0.0001$) (Fig. 7C, open circles). Thus, we confirmed in the population activity the high-precision parametric coding of positive and negative RPE in dopamine neurons.

The distributions of the onset and offset latencies of the CS and reward responses of striatal and dopamine neurons coding RPE are shown in Figure 8. Comparing the distribution of the onset latencies of CS and reward responses between RPE-coding striatal and dopamine neurons (Fig. 8, compare A, E, and C, G), we found no significant differences between them (both $p > 0.1$, Kolmogorov–Smirnov test). Comparing the distribution of the offset latencies of CS and reward responses between RPE-coding striatal neurons and RPE-coding dopamine neurons (Fig. 8, compare B, F, and D, H), we found that offset latencies of CS and reward responses of RPE-coding striatal neurons were longer than those of RPE-coding dopamine neurons ($p < 0.001$ for CS responses and $p < 0.01$ for reward responses, Kolmogorov–Smirnov test).

Invariance of the RPE coding over the change of motivational level

The probability-dependent spout-licking responses were, in all subjects, less frequent in the second half of a daily session than in the first half, and in the majority of subjects, they became so infrequent in all reward probability condition as to be probability nondependent. We think that

the decrease of spout-licking responses and their loss of reward probability dependency reflect the decrease in motivational level attributable to increasing satiety over trials. To examine whether the activities of RPE-coding striatal and dopamine neurons are influenced by the change of the motivational level, we divided the neuron population into two groups, those recorded during the first and second halves of the daily sessions.

Of all 984 neurons recorded in the striatum, 402 neurons were recorded during the first half of a daily session and 582 were recorded during the second half of a daily session. Twenty-two of those 402 and 25 of those 582 were found to code RPE signals. Thus, the chances of finding the RPE-coding striatal neurons were 5.5% (22 of 402) and 4.3% (25 of 582) during the first and the second halves of a daily session, respectively. The result of χ^2 test did not indicate a significant difference between these chances ($p > 0.1$). Of the 51 putative dopamine neurons, 23 were recorded during the first half of a daily session and 28 were recorded during the second half of a daily session. Thirteen of those

23 and 14 of those 28 were found to code RPE signals. Thus, the chances of finding the RPE-coding dopamine neurons were 57% (13 of 23) and 50% (14 of 28) during the first and second halves of a daily session, respectively. The result of the χ^2 test did not indicate a significant difference between these chances ($p > 0.1$).

We plotted the average absolute discharge rate in the CS and reward periods (both within 500 ms after onset) against reward probabilities separately for neurons recorded during the first and the second halves of a daily session, and we did this for both RPE-coding striatal neurons (supplemental Fig. S3A–D, available at www.jneurosci.org as supplemental material) and RPE-coding dopamine neurons (supplemental Fig. S4A–D, available at www.jneurosci.org as supplemental material). Comparing the average absolute discharge rate by two-way ANOVA (the first/second half of a daily session by reward probabilities) revealed no difference between the first and second halves of a daily session in either the CS or reward period activity of RPE-coding striatal and dopamine neurons ($p > 0.1$). We then produced frequency histograms of the slope values of the regression lines between the absolute discharge rate and reward probability in the CS and reward periods for both RPE-coding striatal neurons (supplemental Fig. S3E–H, available at www.jneurosci.org as supplemental material) and RPE-coding dopamine neurons (supplemental Fig. S4E–H, available at www.jneurosci.org as supplemental material). Comparing the distribution of slope values by Kolmogorov–Smirnov test revealed no difference between the first and second halves of a daily session in either the CS or reward period activity of RPE-coding striatal and dopamine neurons ($p > 0.1$). Thus, for neither striatal nor dopamine neurons did we find that the proportion of RPE coding neurons or their absolute discharge rate or sensitivity to probability difference differed between the first and second halves of a daily session. This indicates that the activities of RPE-coding striatal and dopamine neurons are invariant over changes in the motivational level.

Discussion

In this study, we recorded single-unit activity during the performance of a probabilistic pavlovian conditioning task and found that a group of striatal neurons showed activity almost identical with that of midbrain dopamine neurons. In these striatal neurons, the CS and reward responses showed high positive and negative correlations with reward probability, respectively, indicating that these neurons code positive RPE parametrically. To our knowledge, this study is the first to find parametric coding of positive RPE in the noninvasive fashion (i.e., excitatory responses

Figure 7. Baseline-subtracted and peak-normalized population activity of RPE-coding dopamine neurons. **A**, Population histograms of RPE-coding dopamine neurons ($N = 27$) around the CS presentation (left) and the timing of the reward delivery (top right) or omission (bottom right). The lines in different colors show the neuronal activity recorded under different reward probabilities and that at the delivery of unpredicted rewards. **B**, Baseline-subtracted and peak-normalized CS responses of RPE-coding dopamine neurons within 500 ms after CS onset plotted against the reward probability. The CS response shows a positive correlation with the reward probability ($p < 0.0001$; $r = 0.60$). **C**, Baseline-subtracted and peak-normalized reward responses of RPE-coding dopamine neurons at the reward delivery (filled circles) and omission (open circles) plotted against the reward probability. The time window for the response to reward delivery was the first 500 ms after the reward delivery, and that for the response to reward omission was to the first 2000 ms after the reward omission. The reward response shows a negative correlation with the reward probability ($p < 0.0001$; $r = -0.59$). In the 0% reward probability condition, the response to an unpredicted reward was used as a substitute for the reward response. The activity after the time of the reward omission shows a negative correlation with the reward probability ($p < 0.0001$; $r = -0.44$). Error bars show SEM.

at the occurrence of positive RPEs) in neurons other than mid-brain dopamine neurons.

Advantage of electrophysiological recordings in head-fixed rats

We think that the new findings in this study are mostly attributable to the advantages of the behavioral paradigm we used. To evaluate RPE quantitatively, we need to accurately infer what the subject is expecting at a given moment. We therefore need to control the subject and its environment precisely. Head-fixed rats well trained in a probabilistic pavlovian conditioning task would be good subjects for such experimental purposes. Data obtained in freely moving subjects would be hard to interpret because the value of the CS and reward would vary considerably depending on the distance between the subject and the food receptacle. For example, if a rat were near the food receptacle when a CS predicted a pellet of food, the CS would presumably have a higher

Figure 8. Distributions of the onset and offset latencies of the CS and reward responses of RPE-coding striatal neurons ($N = 47$) and RPE-coding dopamine neurons ($N = 27$). Histograms show the proportion of neurons in each time window, and the lines show the cumulative proportion of neurons. **A**, Onset latency of the CS response of RPE-coding striatal neurons. **B**, Offset latency of the CS response of RPE-coding striatal neurons. **C**, Onset latency of the reward response of RPE-coding striatal neurons. **D**, Offset latency of the reward response of RPE-coding striatal neurons. **E**, Onset latency of the CS response of RPE-coding dopamine neurons. **F**, Offset latency of the CS response of RPE-coding dopamine neurons. **G**, Onset latency of the reward response of RPE-coding dopamine neurons. **H**, Offset latency of the reward response of RPE-coding dopamine neurons. The filled bars in **A**, **B**, **E**, and **F** indicate the proportion of neurons showing greater response to the CS indicating 100% reward probability than to the CS predicting no reward, and the open bars in **A**, **B**, **E**, and **F** indicate the proportion of neurons showing no difference between the responses to those CSs. The filled bars in **C**, **D**, **G**, and **H** indicate the proportion of neurons showing greater response to the unpredicted reward than to the reward in the 100% reward probability condition.

value because of temporal discounting (Kobayashi and Schultz, 2008) than it would if the rat were on the other side of the cage. Likewise, the expectation of a food pellet might be higher in those trials, and the response of dopamine neurons to the food pellet might be smaller, because of the short CS–reward interval (Fiorillo et al., 2008). As head fixation is used for most of the electrophysiological recordings in monkeys, its use in rodent studies will fill in the gap between the electrophysiological data obtained from monkeys and rodents. We think that bridging the interspecies gap between monkeys and rodents is extremely important in neuroscience because, on one hand, monkeys are the

closest model to humans and, on the other hand, most of the molecular and other up-to-date neuroscience techniques are applied mainly to rodents.

Common and different features of the activity of RPE-coding striatal and dopamine neurons

We found that the activity patterns of the RPE-coding striatal and dopamine neurons are similar in that both show phasic responses to CSs and rewards with short-onset latencies and both show parametric coding of positive RPE. There was no difference in the distributions of the onset latencies of the CS and reward responses of RPE-coding striatal and dopamine neurons. These may imply that the RPE-coding striatal neurons do not simply reflect the input from dopamine neurons but are actively involved in the coding of the RPE signal by cooperating with dopamine neurons. (The possible neuronal circuit connecting RPE-coding striatal and dopamine neurons is discussed in the next section of Discussion.)

RPE-coding striatal and dopamine neurons are also similar in that the activity of neither group changed when motivational level did over a daily session. When a subject is highly motivated, the reward and the associated CS would be valued more than when the subject is less motivated. Regardless of the CS and the reward value decreases caused by the gradual increase in satiety over the day, the magnitude of the CS and reward responses of RPE-coding striatal and dopamine neurons did not change. This suggests that both groups of neurons adapted to the change of motivational level. In a previous study (Tobler et al., 2005), dopamine neurons were found to show even more drastic adaptation, such as to code the reward value according to the range of available rewards at the time: when medium or large amount of reward was available, dopamine neurons responded inhibitory to medium reward delivery, and when medium or small amount of reward was available, dopamine neurons responded excitatory to medium reward delivery.

The kinds of adaptive coding of reward value observed in striatal and dopamine neurons in this study and in dopamine neurons by Tobler et al. (2005) may be useful in keeping the optimal performance to obtain the available reward at any given times, regardless of the difference of the value of available rewards brought by internal or external factors.

In addition to the many similarities of the activity patterns of the RPE-coding striatal neurons and the RPE-coding dopamine neurons, we also found some differences between them. One was that only the dopamine neurons showed negative RPE coding (i.e., an inhibitory response to the unexpected omission of re-

ward). In the extracellular recording of neurons showing low spontaneous activity, however, inhibitory activity is generally harder to detect than excitatory activity. Confirming the absence of negative RPE coding in RPE-coding striatal neurons will therefore require additional investigations.

Another difference was that the offset latencies of the CS and reward responses were longer in the RPE-coding striatal neurons. This might mean that the RPE information is maintained longer in the striatum than in the midbrain dopamine neurons, which would suggest that there is a functional difference between them in coding RPE signals.

Functional links for coding RPE signals between subcortical structures

The striatum contains various types of neurons, but for the following reasons we think the neurons whose activity we recorded were medium spiny neurons. First, their waveforms were similar to those of typical medium spiny neurons described in the previous studies and were much shorter than those of typical giant aspiny interneurons (Apicella, 2002), which may be the second largest population of striatal neurons. Second, the recorded neurons showed low spontaneous activity and showed phasic excitatory responses to the rewards and CSs, and these characteristics are generally thought to be typical of medium spiny neurons (Apicella, 2007).

Hikosaka and colleagues recently found an inverse RPE coding (i.e., excitatory responses to unpredicted reward omissions and to the CSs predicting no reward) in the GPi (Hong and Hikosaka, 2008) and lateral habenula (Matsumoto and Hikosaka, 2007). The inverse RPE coding in the GPi is in accordance with the result of the present study, as the RPE-coding striatal neurons may be providing inhibitory input to GPi neurons [striatal medium spiny neurons send GABAergic projections to GPi neurons (Bolam et al., 2000)]. The neural circuit connecting RPE-coding striatal neurons with dopamine neurons through GPi, lateral habenula, and RMTg (Herkenham and Nauta, 1979; Bolam and Smith, 1992; Bolam et al., 2000; Jhou et al., 2009a,b) may play a critical role in coding RPE signals. In addition, striatal medium spiny neurons and midbrain dopamine neurons have strong anatomical connectivity through their direct reciprocal projections (Anden et al., 1964; Gerfen, 1985), and there are indirect connections from the striatum to dopamine neurons via the SNr (Hajós and Greenfield, 1994; Tepper et al., 1995; Bolam et al., 2000). Those circuits might also be important for coding RPE signals.

Comparisons of the observed firing properties of the dopamine neurons with those reported in previous studies

Concerning the activity of dopamine neurons, the result of the present study is basically in good accordance with the previous study (Fiorillo et al., 2003) in that the neurons showed parametric modulation of phasic excitatory response to positive RPE that occurred at the CS and reward presentation in a probabilistic pavlovian conditioning. The duration of inhibitory activity we found when the reward was unexpectedly omitted, however, differs from that reported previously. In our study, the inhibitory activity typically lasted for 2 s after the time of reward omission, whereas in their study it lasted for ~200 ms after the time of reward omission. One possible explanation for this difference is that the duration of the inhibitory response to the negative RPE is increased by the ambiguity of the timing of the expected reward. In fact, in the previous studies, the duration of the inhibitory response to the negative RPE appears to be ~200–300 ms when

the timing of the expected reward is made clear by an explicit timing cue (Fiorillo et al., 2003; Satoh et al., 2003; Pan et al., 2008; Bromberg-Martin and Hikosaka, 2009) but appears to be longer when the timing of the expected reward is made unclear by a delay between the timing cue and the reward onset (Hollerman and Schultz, 1998; Morris et al., 2004). However, the ambiguity of the timing of the expected reward does not appear to be the only factor that elongates the duration of the inhibitory response to negative RPE; there has been a study reporting longer duration of inhibitory response to reward omission even with a precise timing cue (Matsumoto and Hikosaka, 2007). Additional study may be needed to identify the factor(s) determining the duration of a dopamine neuron's inhibitory response to negative RPE.

Another difference from the previous studies was that we did not find uncertainty coding in any dopamine neuron. Fiorillo et al. (2003) found that ~30% of the recorded dopamine neurons showed uncertainty-related tonic activity between the CS onset and reward onset. The absence of uncertainty-related activity may be attributable to the delay that we introduced between the CS offset and the reward onset (trace conditioning procedure). Fiorillo et al. (2003) did not have any delay between the CS offset and the reward onset (delay conditioning procedure) in regular recording sessions. They report that introducing a delay between the CS offset and reward onset significantly reduced the uncertainty-related activity preceding the timing of the expected reward onset.

Future perspectives

Not only the midbrain dopamine neurons and the subpopulation of striatal neurons that we have reported but also subcortical neurons in the GPi (Hong and Hikosaka, 2008) and lateral habenula (Matsumoto and Hikosaka, 2007) as well as neurons in the cingulate cortex (Matsumoto et al., 2007; Seo and Lee, 2007) have been found to code RPE signals. Investigating the functional significance of those RPE-coding neurons in different brain areas is important if we are to find out how the RPE signal is generated and used in the brain. To identify the functional role of the RPE-coding striatal neurons, one needs to know how those neurons are connected to the neural networks within and outside of the striatum. It is known that there are different types of medium spiny neurons having different neurotransmitter receptor expressions and different forms of synaptic connectivity with other neurons (Gerfen et al., 1990; Surmeier et al., 2007). Regarding the relatively low proportion of RPE-coding neurons in the whole population of striatal medium spiny neurons, it appears that the RPE-coding striatal neurons are a subset of medium spiny neurons. Future studies involving single-unit recordings during the iontophoretic application of receptor agonists and antagonists, and those involving juxtacellular recordings made under conditions using the behavioral paradigm we used in this study, may reveal the relationships between the activity of RPE-coding striatal neurons and their pharmacological and morphological properties, including receptor expressions and target of axon projections.

References

- Alexander GE, Crutcher MD (1990) Preparation for movement: neural representations of intended direction in three motor areas of the monkey. *J Neurophysiol* 64:133–150.
- Anden NE, Carlsson A, Dahlstroem A, Fuxe K, Hillarp NA, Larsson K (1964) Demonstration and mapping out of nigro-neostriatal dopamine neurons. *Life Sci* 3:523–530.
- Apicella P (2002) Tonically active neurons in the primate striatum and their

- role in the processing of information about motivationally relevant events. *Eur J Neurosci* 16:2017–2026.
- Apicella P (2007) Leading tonically active neurons of the striatum from reward detection to context recognition. *Trends Neurosci* 30:299–306.
- Apicella P, Deffains M, Ravel S, Legallet E (2009) Tonically active neurons in the striatum differentiate between delivery and omission of expected reward in a probabilistic task context. *Eur J Neurosci* 30:515–526.
- Berns GS, McClure SM, Pagnoni G, Montague PR (2001) Predictability modulates human brain response to reward. *J Neurosci* 21:2793–2798.
- Bolam JP, Smith Y (1992) The striatum and the globus pallidus send convergent synaptic inputs onto single cells in the entopeduncular nucleus of the rat: a double anterograde labelling study combined with postembedding immunocytochemistry for GABA. *J Comp Neurol* 321:456–476.
- Bolam JP, Hanley JJ, Booth PA, Bevan MD (2000) Synaptic organisation of the basal ganglia. *J Anat* 196:527–542.
- Bromberg-Martin ES, Hikosaka O (2009) Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron* 63:119–126.
- Canales JJ, Capper-Loup C, Hu D, Choe ES, Upadhyay U, Graybiel AM (2002) Shifts in striatal responsivity evoked by chronic stimulation of dopamine and glutamate systems. *Brain* 125:2353–2363.
- Doron NN, Ledoux JE, Semple MN (2002) Redefining the tonotopic core of rat auditory cortex: physiological evidence for a posterior field. *J Comp Neurol* 453:345–360.
- Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299:1898–1902.
- Fiorillo CD, Newsome WT, Schultz W (2008) The temporal precision of reward prediction in dopamine neurons. *Nat Neurosci* 11:966–973.
- Gerfen CR (1985) The neostriatal mosaic. I. Compartmental organization of projections from the striatum to the substantia nigra in the rat. *J Comp Neurol* 236:454–476.
- Gerfen CR, Engber TM, Mahan LC, Susel Z, Chase TN, Monsma FJ Jr, Sibley DR (1990) D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science* 250:1429–1432.
- Hajós M, Greenfield SA (1994) Synaptic connections between pars compacta and pars reticulata neurones: electrophysiological evidence for functional modules within the substantia nigra. *Brain Res* 660:216–224.
- Haruno M, Kawato M (2006) Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *J Neurophysiol* 95:948–959.
- Herkenham M, Nauta WJ (1979) Efferent connections of the habenular nuclei in the rat. *J Comp Neurol* 187:19–47.
- Hikosaka O, Sakamoto M, Usui S (1989) Functional properties of monkey caudate neurons. I. Activities related to saccadic eye movements. *J Neurophysiol* 61:780–798.
- Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat Neurosci* 1:304–309.
- Hong S, Hikosaka O (2008) The globus pallidus sends reward-related signals to the lateral habenula. *Neuron* 60:720–729.
- Jhou TC, Fields HL, Baxter MG, Saper CB, Holland PC (2009a) The rostromedial tegmental nucleus (RMTg), a GABAergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses. *Neuron* 61:786–800.
- Jhou TC, Geisler S, Marinelli M, Degarmo BA, Zahm DS (2009b) The mesopontine rostromedial tegmental nucleus: a structure targeted by the lateral habenula that projects to the ventral tegmental area of Tsai and substantia nigra compacta. *J Comp Neurol* 513:566–596.
- Kim H, Sul JH, Huh N, Lee D, Jung MW (2009) Role of striatum in updating values of chosen actions. *J Neurosci* 29:14701–14712.
- Kobayashi S, Schultz W (2008) Influence of reward delays on responses of dopamine neurons. *J Neurosci* 28:7837–7846.
- Matsumoto M, Hikosaka O (2007) Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* 447:1111–1115.
- Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007) Medial prefrontal cell activity signaling prediction errors of action values. *Nat Neurosci* 10:647–656.
- McClure SM, Berns GS, Montague PR (2003) Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38:339–346.
- Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H (2004) Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron* 43:133–143.
- O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003) Temporal difference models and reward-related learning in the human brain. *Neuron* 38:329–337.
- 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.
- Pan WX, Schmidt R, Wickens JR, Hyland BI (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 WX, Schmidt R, Wickens JR, Hyland BI (2008) Tripartite mechanism of extinction suggested by dopamine neuron activity and temporal difference model. *J Neurosci* 28:9619–9631.
- Paxinos G, Watson C (2005) The rat brain in stereotaxic coordinates. San Diego: Academic.
- Reynolds JN, Wickens JR (2002) Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw* 15:507–521.
- Reynolds JN, Hyland BI, Wickens JR (2001) A cellular mechanism of reward-related learning. *Nature* 413:67–70.
- Roesch MR, Calu DJ, Schoenbaum G (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat Neurosci* 10:1615–1624.
- Roesch MR, Singh T, Brown PL, Mullins SE, Schoenbaum G (2009) Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. *J Neurosci* 29:13365–13376.
- Rolls ET, Thorpe SJ, Maddison SP (1983) Responses of striatal neurons in the behaving monkey. 1. Head of the caudate nucleus. *Behav Brain Res* 7:179–210.
- Satoh T, Nakai S, Sato T, Kimura M (2003) Correlated coding of motivation and outcome of decision by dopamine neurons. *J Neurosci* 23:9913–9923.
- Schultz W (2002) Getting formal with dopamine and reward. *Neuron* 36:241–263.
- Schultz W (2006) Behavioral theories and the neurophysiology of reward. *Annu Rev Psychol* 57:87–115.
- Seo H, Lee D (2007) Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J Neurosci* 27:8366–8377.
- Surmeier DJ, Ding J, Day M, Wang Z, Shen W (2007) D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. *Trends Neurosci* 30:228–235.
- Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT.
- Tepper JM, Martin LP, Anderson DR (1995) GABA_A receptor-mediated inhibition of rat substantia nigra dopaminergic neurons by pars reticulata projection neurons. *J Neurosci* 15:3092–3103.
- Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. *Science* 307:1642–1645.
- Wickens JR, Begg AJ, Arbuthnott GW (1996) Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. *Neuroscience* 70:1–5.