Different Neural Correlates of Reward Expectation and Reward Expectation Error in the Putamen and Caudate Nucleus During Stimulus-Action-Reward Association Learning

Masahiko Haruno and Mitsuo Kawato

JN 95:948-959, 2006. First published Sep 28, 2005; doi:10.1152/jn.00382.2005

J Neurophysiol 95: 948–959, 2006.
First published October 5, 2005; doi:10.1152/jn.00382.2005.

Different Neural Correlates of Reward Expectation and Reward Expectation Error in the Putamen and Caudate Nucleus During Stimulus-Action-Reward Association Learning

Masahiko Haruno[1,2] and Mitsuo Kawato[1]

[1]*Advanced Telecommunication Research Institute Computational Neuroscience Laboratories, Department of Cognitive Neuroscience, Kyoto, Japan; and* [2]*Institute of Neurology, University College London, London, United Kingdom*

Submitted 14 April 2005; accepted in final form 26 September 2005

Haruno, Masahiko and Mitsuo Kawato. 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, 2006. First published October 5, 2005; doi:10.1152/jn.00382.2005. To select appropriate behaviors leading to rewards, the brain needs to learn associations among sensory stimuli, selected behaviors, and rewards. Recent imaging and neural-recording studies have revealed that the dorsal striatum plays an important role in learning such stimulus-action-reward associations. However, the putamen and caudate nucleus are embedded in distinct cortico-striatal loop circuits, predominantly connected to motor-related cerebral cortical areas and frontal association areas, respectively. This difference in their cortical connections suggests that the putamen and caudate nucleus are engaged in different functional aspects of stimulus-action-reward association learning. To determine whether this is the case, we conducted an event-related and computational model–based functional MRI (fMRI) study with a stochastic decision-making task in which a stimulus-action-reward association must be learned. A simple reinforcement learning model not only reproduced the subject's action selections reasonably well but also allowed us to quantitatively estimate each subject's temporal profiles of stimulus-action-reward association and reward-prediction error during learning trials. These two internal representations were used in the fMRI correlation analysis. The results revealed that neural correlates of the stimulus-action-reward association reside in the putamen, whereas a correlation with reward-prediction error was found largely in the caudate nucleus and ventral striatum. These nonuniform spatiotemporal distributions of neural correlates within the dorsal striatum were maintained consistently at various levels of task difficulty, suggesting a functional difference in the dorsal striatum between the putamen and caudate nucleus during stimulus-action-reward association learning.

INTRODUCTION

Because learning appropriate behaviors for given situations through reward and penalty information is crucial for living, the neural mechanisms involved in reward-based behavioral learning have attracted enormous attention in system neuroscience. When we are confronted with new and uncertain situations, acquisition of stimulus-action-reward association seems essential for selecting the optimal behavior. This is because a better action for a given situation can be easily selected by comparing expected rewards according to the combination of a specific contextual cue, which characterizes the situation and possible choices of action.

Previous human and nonhuman primate studies have indicated that the dorsal striatum is a key brain structure for learning such prediction. When human subjects learn to select appropriate behaviors in a stimulus-action-reward association task, the caudate activity in each learning block is correlated with the amount of behavioral change that the subject makes in a block (Haruno et al. 2004). Another functional MRI (fMRI) study reported that activity in the anterior striatum (mainly the caudate nucleus) is correlated with the reward-prediction (TD) error during behavioral learning (O'Doherty et al. 2004). A considerable number of fMRI studies have also revealed a correlation between activity in the ventral striatum and reward-prediction error in regard to both primary rewards (Berns et al. 2001; McClure et al. 2003; O'Doherty et al. 2003; Pagnoni et al. 2002) and monetary rewards (Breiter et al. 2001; Knutson et al. 2001). In addition, caudate activity is reported to be correlated with prediction of reward in tasks that do not include behavioral learning (i.e., stimulus-reward association) (Delgado et al. 2000; Tricomi et al. 2004). Consistent with human data, neural-recording studies on monkeys have shown the involvement of the putamen (Hikosaka et al. 1999; Matsumoto et al. 1999; Tremblay et al. 1998) and caudate nucleus (Kawagoe et al. 2001; Shidara et al. 1998; Tremblay et al. 1998) in reward association learning tasks.

Closely related to learning in the striatum, midbrain dopamine neurons projecting to the striatum fire at reward delivery before learning, while the activity shifts forward in time to the presentation of a reward cue when the reward is predictable from the cue (Hollerman and Schultz 1998; Schultz et al. 1992, 2003; Takikawa et al. 2004). The reinforcement learning models can explain this temporal shift (Schultz et al. 1992) in terms of the temporal difference (TD) error, suggesting that reward prediction, whether action-dependent or action-independent, is learned in the dorsal striatum by using the TD error (Brown et al. 1999; Houk et al. 1995; Montague et al. 1996). There are several possible implementations of TD models, but the two most prevalent examples are the actor-critic architecture and Q-learning (Sutton and Barto 1998). The former learns the action-independent evaluation of context (critic) and how to act in the context (actor) separately, whereas the latter acquires a single representation of stimulus-action-reward association

Address for reprint requests and other correspondence: M. Haruno, Department of Cognitive Neuroscience Computational Neuroscience Labs, Advanced Telecommunication Research Institute, 2-2-2 Hikaridai Seikacho, Sorakugun Kyoto 619-0288, Japan (E-mail: mharuno@atr.jp).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

dubbed the Q-table. All of the experimental data described above are consistent with the TD models.

Nevertheless, no previous experimental or modeling study has incorporated the following anatomical findings, which could enhance our functional understanding of the dorsal striatum during stimulus-action-reward association learning. Specifically, the anatomical connections of the putamen are dominant in sensory-motor-related areas such as the premotor and primary motor cortices (Alexander et al. 1990; Gerardin et al. 2003; Parthsarathy et al. 1992; Selemon and Goldman-Rakic 1985; Takada et al. 1998), while those of the caudate nucleus are dominant in sensory-reward-related areas such as the orbitofrontal and prefrontal cortices (Alexander et al. 1990). This difference suggests that the putamen is involved mainly in evaluating actions in terms of sensory contexts and rewards, whereas the caudate nucleus is involved mainly in comparing actual and predicted rewards for learning.

To empirically examine this hypothesis, we conducted an fMRI study of a stimulus-action-reward association task in which subjects were asked to learn an advantageous button-push (left or right) in response to visual stimuli. In this task, the combination of the stimulus and the subject's button push stochastically determines monetary reward. The visual stimulus, button-push, and delivery of monetary reward cue were separated from each other, by several to 10 s, in this order. From the temporal design of the task, we could expect that subjects needed to form stimulus-action-reward association for decision-making at the stimulus onset, whereas the comparison between actual and predicted rewards could be made only at reward delivery timing. To estimate the stimulus-action-dependent reward prediction (SADRP) and reward-prediction error (RPE) in each trial within the subjects' brains, we adopted the Q-learning model because it handles stimulus-action-reward association directly. Importantly, RPE in this study is not identical to the TD error. Within the context of this study, the relationship between TD error and our SADRP and RPE is as follows. At the beginning of learning, RPE is nearly equivalent to TD error, and SADRP is close to 0. At the later stage of learning, RPE at the reward delivery timing corresponds to the stochastic component of TD error, because the predictable reward can already be estimated by SADRP at the cue timing. Accordingly, the temporal difference of SADRP at the cue timing is equivalent to the predictable part of the TD error. However, the temporal dissociation between SADRP and RPE is critical to our hypothesis on the different contributions of the putamen and caudate nucleus. Therefore we conducted an event-related correlation analysis of fMRI data with SADRP and RPE.

METHODS

Subjects

Twenty healthy adults (23–31 yr old, 11 males and 9 females, all right-handed) participated in the experiment. Informed consent of the participants was obtained beforehand, and the protocol was approved by the institution's ethics committee.

Experimental design

In a TEST trial (Fig. 1*A*), subjects learned the stochastic association between a visual stimulus, a button-push, and rewards to maximize their total monetary rewards. In each trial, after one of three fractal stimuli (FSs) was presented (onset at 0.7 s), subjects pushed the left or right button following a beep sound (randomized at 5.2 or 6.2 s). A small green circle appeared either to the left or right of the fixation cross to show which button had been selected. All subjects pushed the buttons with the index or middle finger of their right hand. If the trial was successful, the figure frame turned yellow (randomized at 10.2 or 12.2 s) and the subject earned a 50-yen reward. Otherwise, the frame turned purple and the subject suffered a 50-yen penalty (not shown in Fig. 1*A*).

The actual outcome of each button-push—success or failure—was stochastically controlled depending on the fractal stimulus presented and the subject's button-push. As an example of how this stochasticity works, Fig. 1 shows experimental session 2 (S2 out of three sessions, S1–S3), which was controlled with a probability of 0.8 (80%). For the purple fractal figure (FS1) in this example, a left button-push yielded +50 yen with a probability of 0.8 and −50 yen with a probability of 0.2. A right button-push, on the other hand, yielded +50 yen with a probability of 0.2 and −50 yen with a probability of 0.8. Therefore the optimal behavior for FS1 was to push the left button, which the subjects had to learn by trial and error. In S2, the dominant probabilities of the other two fractal figures (FS2 and FS3) were also 0.8, and the advantageous button-push was randomized for left or right (optimal behaviors were FS1: left, FS2: right, and FS3: left). Note that subjects could not develop a stimulus-action-reward association before presentation of the FS. Importantly, the subjects were instructed to decide which button to push as soon as the FS was presented. This suggests that the subjects most likely associated action, stimulus, and reward for decision-making at the time of stimulus presentation (shown as SADRP in Fig. 1*A*) and computed the reward prediction error at the time of reward delivery (shown as RPE in Fig. 1*A*), which would validate the event-related fMRI analysis later. The occurrence of the three fractal figures was controlled equally and pseudorandomly by using the same random number sequence for all subjects to reduce the variance of learning speed across subjects. Each trial lasted 19 or 21 s, and one TEST block included four repetitions of a trial (Fig. 1*C*). The accumulated reward was displayed above the figure frame and updated at the moment of reward delivery.

In a CONTROL trial (Fig. 1*B*), the subjects passively pushed the same button as in the preceding TEST block. They were signaled which button to push by a small green circle that appeared to the left or right of the fixation point just after fractal stimulus presentation; this reproduced their own button-push in the preceding TEST block in a randomized order. The fractal stimulus and the outcome color (yellow or purple) had no influence on the subject's button selection but simply reproduced the effects of the visual displays in the TEST trials. Thus aside from the timing of the green circle's presentation, the CONTROL block reproduced all of the physical events of the preceding TEST block and was used to subtract these effects from the TEST trials. No reward or penalty was given in the CONTROL trial. The accumulated reward above the figure box in the CONTROL block remained constant at the value of the preceding TEST trial. As in the TEST block, one trial lasted 19 or 21 s, with four repetitions per block, and the TEST and CONTROL blocks were alternated (Fig. 1*C*). One session included 12 TEST/CONTROL blocks and lasted 32 min [20 s (on average) × 4 trials × 2 (TEST + CONTROL) × 12 blocks]).

We conducted three experimental sessions, S1, S2, and S3, in which the dominant probability was 0.9, 0.8, and 0.7, respectively. According to the stochastic uncertainty, learning was expected to become progressively more difficult. The order of these sessions was counterbalanced across the subjects, and the results were analyzed together because no marked differences in learning performance or imaging results were found. At the start of the experiments, the subjects were told that success or failure depended stochastically on the fractal stimulus presented and the button pushed, but they were not provided with any concrete information on stochastic parameters. The subjects were encouraged to earn as large a monetary reward as

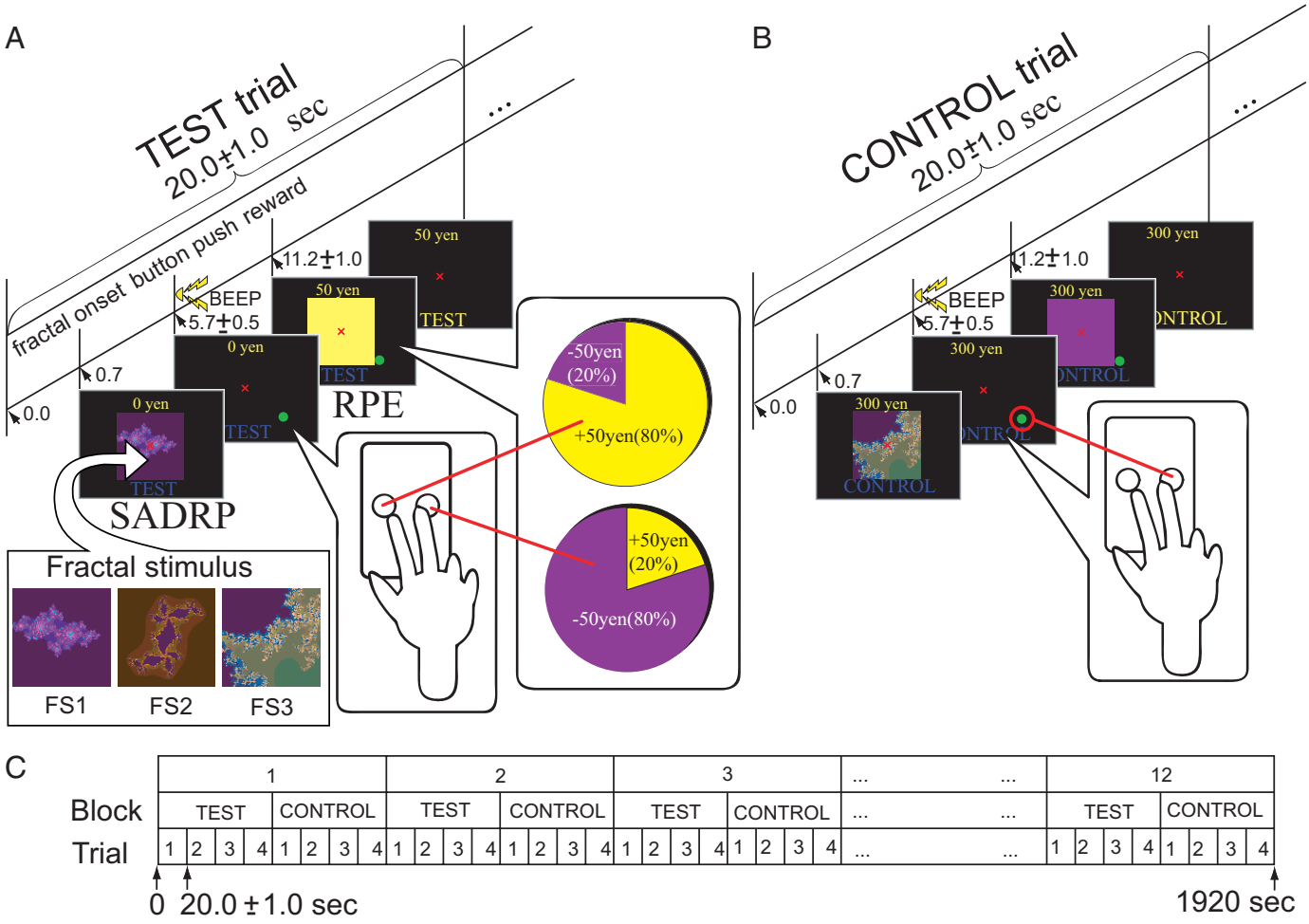

FIG. 1. Experimental design. *A–C*: TEST trial, CONTROL trial, and overall organization of the experiment. *A*: in each TEST trial, 1 of 3 fractal stimuli (FS) was presented, and the subject was required to press the left or right button after a beep to obtain a monetary reward. A small green circle appeared showing which button the subject had pushed. In this example (session 2), the optimal (advantageous) button-push for each FS was set (FS1: left, FS2: right, FS3: left) to yield a reward of 50 yen (yellow frame presented) or a penalty of −50 yen (purple, data not shown) with a probability of 0.8 and 0.2, respectively. In contrast, a nonoptimal (disadvantageous) button-push (FS1: right, FS2: left, FS3: right) led to a 50-yen reward or penalty with a probability of 0.2 and 0.8, respectively. *B*: in CONTROL, subjects had to reproduce their button-pushes in the preceding TEST block for the same set of fractal stimuli while visually instructed by the position of the small green circle. Order of stimulus and button-push was randomized. FS and outcome color (yellow or purple) simply reproduced TEST and was unrelated to the subjects' selection of button-push. No reward or penalty was given in CONTROL, and the accumulated reward above the figure frame remained constant at the value of the preceding TEST trial. *C*: TEST and CONTROL blocks each included 4 trials, and they were interleaved 12 times.

possible, and it was actually given to them in addition to their basic compensation (1,500 yen). We prepared five different sets of three fractal stimuli and changed the configuration of the stimulus set for every session to exclude any brain activity arising from a fixed set of figures.

Computational model for estimating SADRP and RPE

A reinforcement learning model was introduced to estimate the subject's SADRP and RPE during learning. There is a notable difference between SADRP and the conventional "reward prediction" mentioned in previous physiological and imaging studies, in which the reward prediction was a reward amount predicted solely from a given sensory cue but unrelated to actions or selection of behaviors. More precisely, a subject's SADRP at time t can be represented as a table $Q_t(fs, bp)$ indicating the predicted amount of reward for a button-push bp (right or left) and a fractal stimulus fs. Because the number of components is equal to the product of the number of stimuli (3) and the number of actions (2), in this experimental paradigm, SADRP consists of six components. Note that the optimal selection of behaviors is trivial once the true SADRP table is acquired; at that point, the button with the larger Q is selected. When the subject receives an actual reward r_t, the RPE amounts to $r_t - Q_t(fs, bp)$. Then, the model changes the element of the table by the following rule so as to decrease the RPE for the next occurrence of the same combination of stimulus and action

$$Q_{t+1}(fs,bp) = Q_t(fs,bp) + \alpha_t^{fs}[r_t - Q_t(fs,bp)]$$

This procedure, which only updates the table element corresponding to the subject's selected action bp and the given fractal stimulus fs in proportion to the reward prediction error, is known as the "Q-learning algorithm" (Sutton and Barto 1998). It is used here to estimate subjects' SADRP and RPE. Therefore only the component of SADRP that corresponds to the given stimulus and the selected action in each trial will be shown, updated, and used in the subsequent analysis. In the early stage of learning, when SADRP is inaccurate and RPE has a large value, the change in SADRP is expected to be large, whereas in the late stage of learning when SADRP is accurate and RPE is small, the change in SADRP is expected to be small. Thus SADRP tends to converge to an asymptotic value.

The learning rate α_t^{fs} controls the amplitude of change and is determined by a standard recursive least-square procedure (Bertsekas and Tsitsiklis 1996; Dayan et al. 2000; Young 1984). In the current situation, α_t^{fs} reduces to an estimation of the inversed variance for the fractal stimulus *fs* that has a value of 1 when presented and 0 otherwise; then we derive the following update rule

$$\alpha_t^{fs} = \frac{\alpha_{t-1}^{fs}}{1 + \alpha_{t-1}^{fs}}$$

Qualitatively, the learning rate α_t^{fs} decreases as SADRP becomes reliable. This property of α_t^{fs} is important because SADRP does not necessarily change much after the completion of learning, even if RPE occurs because of the stochastic nature of the task. The update equation indicates that the learning rate sharply decreases below 1, suggesting that the initial value of α_t^{fs} (i.e., α_0^{fs}) has little effect on the estimation of SADRP and RPE. We actually examined values of 10, 100, 1,000, 10,000, and 100,000 and confirmed that the resulting SADRP and RPE were not sensitive to them. Therefore we set a value of 1,000 throughout the study.

Finally, it was possible to evaluate the model by examining how often an actual subject's behaviors and advantageous *bp* in terms of $Q_t(fs, bp)$ agreed with each other.

MRI acquisition and preprocessing

MRI scanning was conducted with a 1.5-T Marconi scanner. For each subject, 768 scans of BOLD images (TR 2.5 s, TE 49 ms, flip angle 80°, FOV 192 mm, resolution 3 × 3 × 5 mm) were acquired over two sessions. In addition to these experimental trials, each session contained two preliminary dummy CONTROL trials (16 scans) to allow for T1 equilibration effects. Then, we stopped the MRI scanner and let subjects out for a 10-min break outside the scanner. After the break, the same procedure was repeated for another (3rd) session. High-resolution [T1 (1 × 1 × 1 mm) and T2 (0.75 × 0.75 × 5 mm)] structure images were also acquired for each subject. The data were analyzed using standard procedures implemented in Statistical Parametric Mapping (SPM99) (Friston et al. 1995). Before statistical analysis, we conducted motion correction and nonlinear transformation into the standard space of the MNI coordinates as implemented in SPM99. These normalized EPI images were resliced into 2 × 2 × 2-mm voxels and smoothed with an 8-mm full-width half-maximum isotropic Gaussian kernel.

Computational model–based regression analysis

After preprocessing, we analyzed the data following the standard procedure of the random effect model implemented in SPM99. Specifically, we conducted an event-related correlation analysis of fMRI data with SADRP and RPE. We assumed that brain activities related to SADRP and RPE occur at the timing of the stimulus presentation and reward delivery, respectively. The accuracy of this timing assumption is discussed earlier. SADRP and RPE during the CONTROL trials were assumed to be 0. This assumption is justified as follows. First, there was no monetary reward. Second, the combination of fractal stimuli and button-pushes (left or right) was arbitrary during control trials. Therefore it was neither necessary nor possible for subjects to predict the amount of rewards during CONTROL. Third, the subjects were instructed to push the button passively.

Figure 2 shows how regressors were constructed from SADRP and RPE. The Q-learning model was used to estimate each subject's SADRP and RPE in each trial (Fig. 2*A*). The *i*th and *j*th trials shown here schematically represent the early and late learning phases, respectively. To model the BOLD signal driven by SADRP and RPE, these two variables were convolved with a hemodynamic response function (Fig. 2*B*, spm_hrf function with TR equal to 2.5). The waveforms of the two regressors were determined as shown in Fig. 2*C*

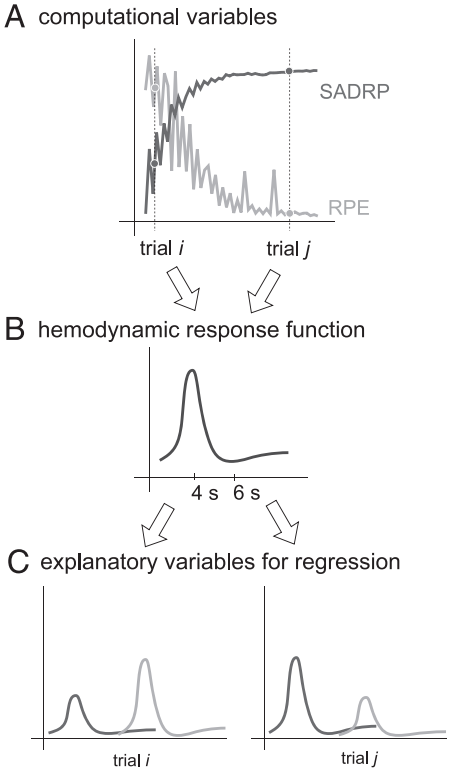

FIG. 2. Regression analysis with stimulus-action-dependent reward prediction (SADRP) and reward-prediction error (RPE). *A*: each subject's SADRP and RPE in each trial were estimated by the Q-learning algorithm. *B*: SADRP and RPE were convolved with a hemodynamic response function to model the BOLD signal representing SADRP and RPE. *C*: resulting time-courses of regressors for SADRP and RPE in the *i*th and *j*th trials.

in each trial, based on the assumption that two brain activities started at the stimulus presentation and at reward feedback. These two regressors do not overlap within a trial as shown in Fig. 2*C*, which helped to make the event-related correlation analysis reliable.

Statistical threshold and illustrations

The statistical threshold was set at $P < 0.001$, uncorrected for multiple comparisons, with the additional constraint that at least five contiguous voxels be included. This uncorrected threshold could be supported because only the striatum was our region of interest. As for the conjunction of ASDRP and RPE over the three sessions (S1–S3) shown in Fig. 9, we simply extracted the voxels with a t value >3.0 in all three sessions by applying a masking operation. We selected this method because we could not directly compare statistics derived from different scanning sessions. We also examined another threshold of I value = 3.5, and the results were quite similar to the case of 3.0. All of the illustrations of statistical maps (i.e., Fig. 6–10) were prepared using our in-house software named "multi_color," which is freely available to the research community (http://www.cns.atr.jp/multi_color/).

RESULTS

Behavioral results

Figures 3–5 show how the reward acquisition and button-push behaviors changed during the TEST blocks of the stimulus-action-reward association task for the most successful subject (Fig. 3) and least successful subject (Fig. 4) in terms of total monetary reward, and the average for the 20 subjects (Fig.

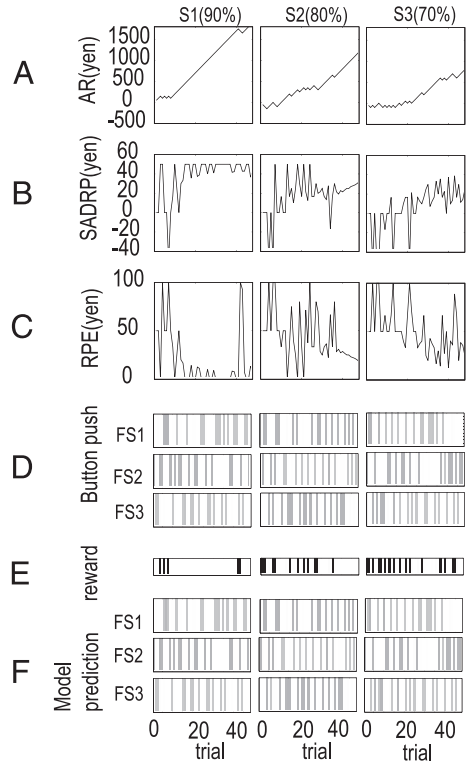

FIG. 3. Behavioral results of learning for the most successful subject in terms of total reward. *A–C*: time-courses of accumulated reward (AR), SADRP, and RPE. *D* and *F*: chronological plots of actual button-pushes by the subjects and corresponding model predictions for each fractal stimulus (FS1–3), respectively, aligned with the subjects' actual rewards (*E*). In *D* and *F*, light grey and dark grey bars represent a left and right button-push, respectively, whereas in *E*, white and black bars represent a reward and penalty, respectively. S1, S2, and S3 represent experimental sessions with a dominant probability of 0.9, 0.8, and 0.7, respectively.

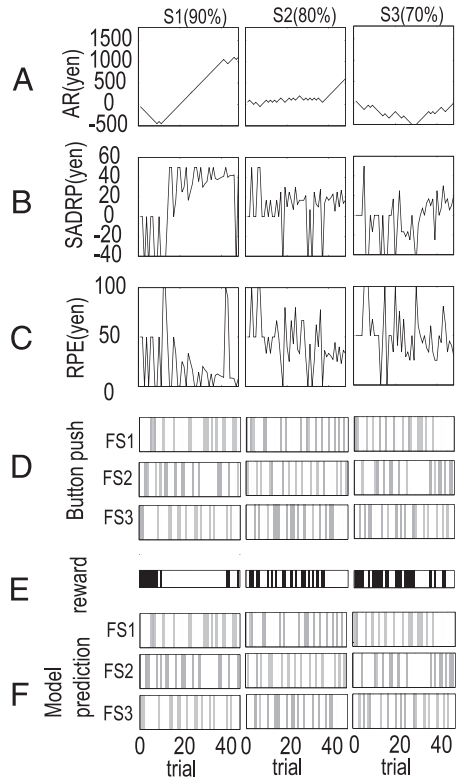

FIG. 4. Behavioral results of learning for the least successful subject in terms of total reward. All subplots follow format of Fig. 3.

5). Accumulated reward (AR) increases almost monotonically in S1–S3 in Fig. 3. In contrast, only S1 exhibits a monotonic increase in Fig. 4, and the flat and decreasing tendencies found in S2 and S3 show that learning was demanding for the subject and that it had not yet been completed within the given number of trials. The averages of all subjects displayed in Fig. 5 show that ARs yielded progressively smaller positive slopes in S1, S2, and S3. Accumulated rewards in the final TEST blocks were significantly larger than zero ($P < 0.0001$; *t*-test) and ranked in the order S1 > S2 > S3 ($P < 0.05$; *t*-test). These observations are consistent with the hypothesis that learning is progressively more difficult in S1, S2, and S3 in accordance with their stochastic uncertainties.

From their behavior, we estimated each subject's SADRP by the Q-learning model (Sutton and Barto 1998), which is defined as the amount of reward predicted by a subject based on a given contextual stimulus and an action selected by the subject. The RPE amounts simply to the difference between SADRP and an actual reward. SADRP is shown in Figs. 3*B*, 4*B*, and 5*B*. The horizontal lines in Fig. 5*B* show theoretical maximum values that are expected for optimal button-push {40 yen [$=50 \times (0.9 - 0.1)$], 30 yen [$=50 \times (0.8 - 0.2)$], and 20 yen [$=50 \times (0.7 - 0.3)$] for S1–S3, respectively}. In the easiest task (S1), SADRP increased and approached the theoretical maximum (40 yen) within 20 trials for all subjects. In more stochastic tasks (S2 and S3), the increase in SADRP became progressively slower than in S1, and some of the subjects failed to achieve the maximum SADRP even in the final TEST trial. None of the estimated SADRPs of any of the subjects showed a simple monotonically increasing tendency

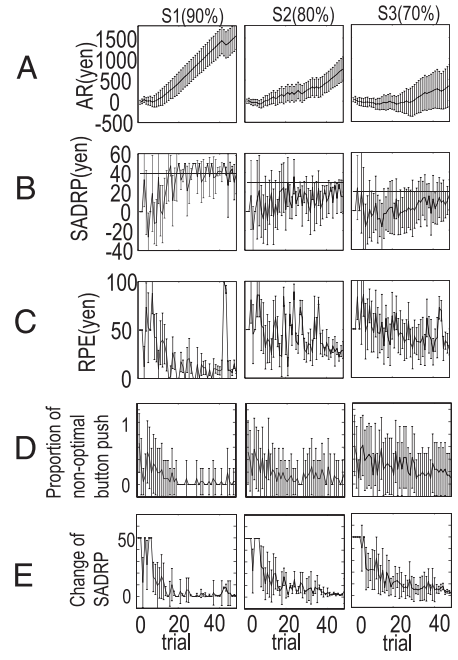

FIG. 5. Behavioral results of learning for the average and SD of all 20 subjects. Corresponding to Figs. 3 and 4, *A–C* show time-courses of AR, SADRP, and RPE averaged over 20 subjects. *D* and *E*: proportion of nonoptimal button-pushes by subjects and change in SADRP of the model, respectively.

because of the stochasticity of the task. Furthermore, it is even difficult to find general increasing tendency in the more stochastic S2 and S3 tasks among the poorer subjects (e.g., Fig. 4). Considering this nonmonotonic nature of SADRP, the subsequent regression analysis of fMRI data with SADRP did not simply capture the artifact correlated with an arbitrary increasing function in time.

Corresponding to SADRP, the absolute values for RPE shown in Figs. 3*C*, 4*C*, and 5*C* quickly decreased to close to 5 yen within 20 trials in S1, but decreased only slowly in S2 and S3. The absolute value was taken because BOLD signal change in the striatum is assumed to represent the energy consumption that arises from the synaptic plasticity change triggered by the RPE. The spiked increase of RPE found in the final stage of S1 (see Figs. 3*C*, 4*C*, and 5*C*) was induced because an unexpected penalty (−50 yen) with low probability occurred, whereas the majority of subjects predicted a 40-yen reward (−50 − 40 = −90 yen RPE). This is also evident in the average (Fig. 5), because most subjects who had already learned to predict a positive reward received an unexpected penalty at this point because of our use of the same random-number sequence. Again, because of the stochasticity of the task, the RPEs did not exhibit a monotonically decreasing tendency in time. It is also difficult to find generally decreasing patterns in the most stochastic S3 tasks among the poorer subjects (e.g., Fig. 4). Thus regression with RPE again did not simply capture brain activity that was correlated with an arbitrary decreasing function in time.

To evaluate how well the simple Q-learning model predicted each subject's behaviors, Figs. 4*D* and 5*D* also compare the actual button-pushes, which subjects selected for each of the fractal stimuli during the TEST trials, and the corresponding behaviors (Figs. 4*F* and 5*F*) predicted by the model. These subject and model behaviors were aligned with the actual reward (Figs. 4*E* and 5*E*), in which a reward and a penalty are labeled in white and black, respectively. In Figs. 4, *D* and *F*, and 5, *D* and *F*, FS1–3 are represented from *top* to *bottom*, with the abscissa showing the number of trials in the temporal order of presentation of the three stimuli. Light grey and dark grey vertical bars represent left and right button-pushes, respectively. In the model, we assumed that each subject's button-push was selected according to which button-push, left or right, was more advantageous in terms of the SADRP table (deterministic selection: the button with the larger Q is always selected).

The model's predictions showed generally good agreement with subjects' actual behaviors. In the most successful subject (Fig. 3, *D* and *F*), the behaviors and predictions were different only in the first few trials, with the discrepancy seeming to arise from a difference in initial strategies, in which the model set the elements of SADRP at 0, thus setting button selection probabilities for left and right equally at 0.5. For the least successful subject (Fig. 4, *D* and *F*), the model's predictions and actual behaviors coincided very well in the easiest task (S1), but the degree of agreement decreased progressively in S2 and S3. This subject's behaviors changed more frequently than the model's prediction. A possible reason for the discrepancy is that the subject was naïve to an unfortunate penalty (see also Fig. 4*E*) because of stochastic uncertainty and behaved in a shortsighted and non–self-confident way without considering the long-term statistics of reward and penalty. This suggests that the subject was more explorative than the behavior expected from using the Q-learning algorithm. Averaged over all 20 subjects, the mean precision of the model's prediction was 0.92 ± 0.21 (SD), 0.85 ± 0.32, and 0.73 ± 0.42 for S1, S2, and S3, respectively. These values indicate that this parsimonious model simulated the subjects' behaviors reasonably well.

Both the simplicity of the model and its ability to predict behaviors motivated the use of computational internal representations such as SADRP and RPE in the subsequent fMRI analysis. In addition, Fig. 5, *D* and *E*, compare the proportion of nonoptimal button-pushes and the change in SADRP averaged over all subjects. This ratio was determined from the subject's behaviors alone. It decreased most rapidly in S1 and progressively more slowly in S2 and S3, reflecting the increasing stochastic uncertainty and resulting greater difficulty. The later stage of the proportion of nonoptimal button-pushes showed smaller fluctuations than later-stage RPE, although the fluctuations decreased in both with the number of trials. The time-course of the change in SADRP showed a pattern of decay closer to that of the proportion of nonoptimal button-pushes than that of RPE, which continuously fluctuated until the end of the learning trials because of the stochastic uncertainty of the task. This contrast shows that the change in SADRP better explains each subject's behavioral learning (the proportion of nonoptimal button-pushes) than RPE does, suggesting that SADRP better reflects the internal representations responsible for behavioral learning. In summary, all of the observations described above indicate that the learning strategy of the human subjects is reasonably comparable with a very simple computational model based on SADRP and RPE.

fMRI results

We carried out an event-related regression analysis of the fMRI data in the striatum with SADRP and RPE. All analyses were conducted with the random-effect model implemented in SPM99 (Friston et al. 1995), and the statistical threshold was set at $P < 0.001$, uncorrected for multiple comparisons, with the additional constraint that at least five contiguous voxels be included. We assumed that the processing related to SADRP and that to RPE are two temporally distinct events triggered by the presentation of the fractal stimulus and by reward delivery, respectively. This assumption is reasonable considering the instruction that subjects should decide on a button-push for a FS at its onset and the fact that there was an interval of >10 s between FS presentation and reward delivery (see also Figs. 1 and 2). In other words, the hemodynamic response for SADRP was assumed to begin to rise on fractal presentation and to reach a peak magnitude proportional to SADRP a few seconds later. Similarly, the hemodynamic response for RPE was assumed to begin to rise on reward delivery and to reach a peak magnitude proportional to RPE. The correlation analyses for the two variables in different sessions (S1–S3) were conducted separately because the scanner was stopped and the subjects went for a 10-min break between their second and third sessions.

Figure 6 shows the correlated activity in the striatum (consisting of the putamen and caudate nucleus) with SADRP and RPE for the simplest task S1 (Fig. 6, *A* and *B*; identical data with a right and left view). Here, the color map associated with each voxel represents its T-values of SPM99 for SADRP and

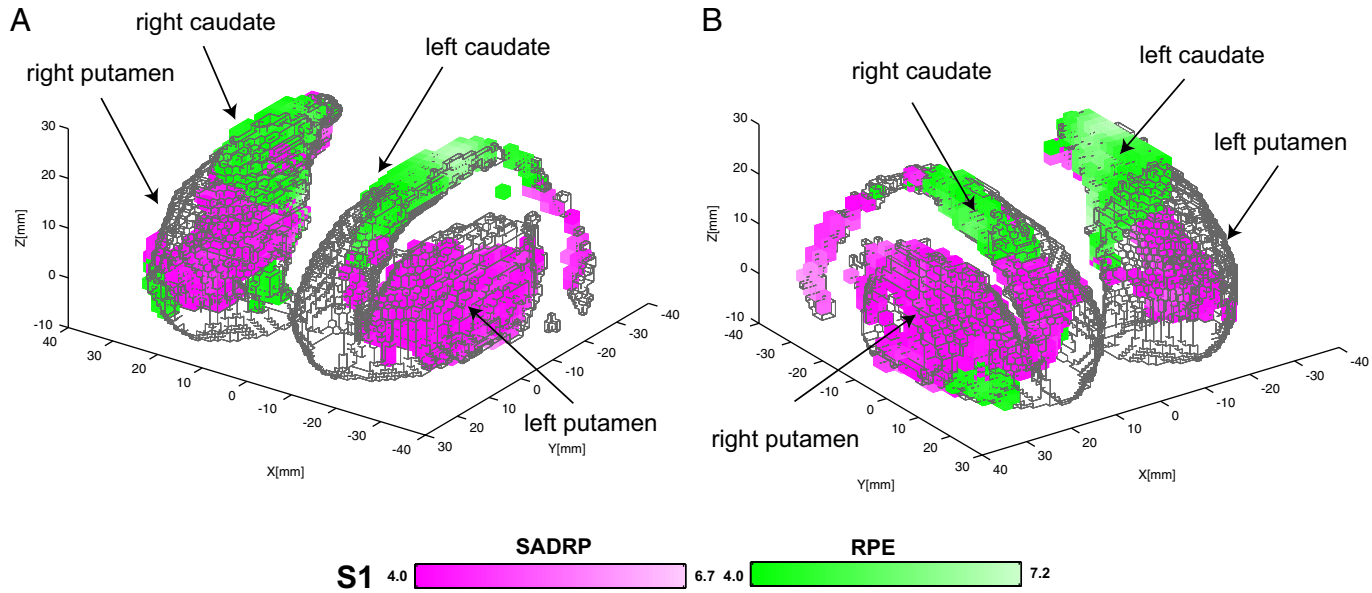

FIG. 6. Activity in the striatum correlated with SADRP and RPE for S1. Each voxel (1 × 1 × 1 mm) is associated with T-values for SADRP (pink) and RPE (green), which are represented as the brightness of the colors as shown in color bars. These activity maps were constructed by reslicing the original activity map (2 × 2 × 2 mm). Overlapping voxel activated in the 2 analyses is represented by a mosaic comprising 2 corresponding colors. Range of MNI coordinates in the illustration is ([−35, 35], [−35, 25], [−8, 24]), which includes the entire dorsal striatum (putamen and caudate nucleus) and part of the ventral striatum (ventral putamen; Talairach and Tournoux 1998). Peak T-values for SADRP activity in the left and right putamen were 6.42 and 5.99; for RPE activity in the left and right caudate nucleus were 9.82 and 5.54, and for the left and right ventral striatum activity were 5.45 and 5.13, respectively.

RPE in pink and green, respectively. The MNI coordinates of the peak activity for SADRP were [−16,−2,0], located at the boundary between the anterior and intermediate putamen in the vicinity of the anterior commissure (Talairach and Tournoux 1998). In contrast, the peak voxel correlated with RPE was at [−8,−4,6], located in the caudate nucleus. A strong correlation with RPE was also found in the ventral striatum, where the MNI coordinates of the peak voxel were [−10,−2,−2]. Both SADRP and RPE activities were bilateral, although T-values for the left-side activity were larger than those for the right-side activity.

For S2 and S3, the striatal activity correlated with SADRP (red and orange, respectively) and RPE (cyan and magenta, respectively) are shown in Figs. 7 and 8 in the same format as in Fig. 6. The peak activity correlated with SADRP and RPE were again found at the boundary between the anterior and intermediate putamen ([−20,−6,4] and [−26,0,4] for S2 and S3, respectively) and in the caudate nucleus ([−12,6,10] and [−12,0,14] for S2 and S3, respectively). These activities were bilateral, and T-values for the left-side activity were slightly larger than those for the right-side activity. The correlation with RPE was also found in the ventral striatum ([−10,0,−4]

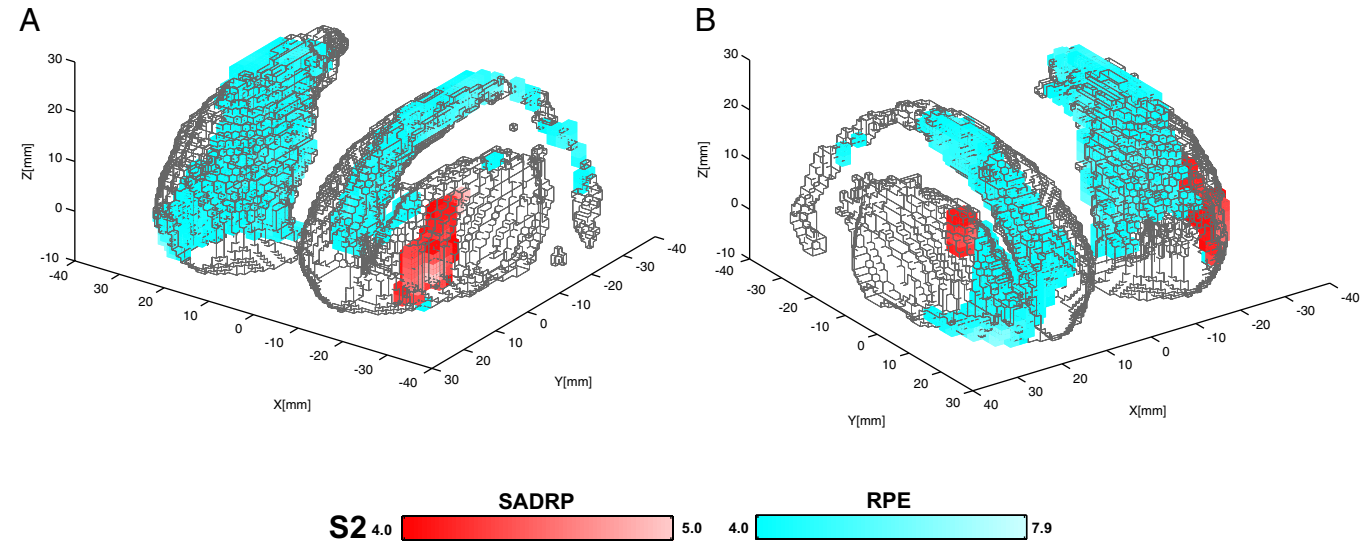

FIG. 7. Activity in the dorsal striatum correlated with SADRP (red) and RPE (cyan) for S2. Figure is in the same format as Fig. 6. Peak T-values for SADRP activity in the left and right putamen were 6.67 and 5.99, for RPE activity in the left and right caudate nucleus were 7.89 and 6.81, and for the left and right ventral striatum activity were 9.31 and 9.29, respectively.

and ([−10,0,−4] were the peaks for S2 and S3, respectively). Overall, the activities for S2 and S3 showed the same tendencies as that for S1. The only notable difference was that the number of correlated voxels with SADRP and RPE became smaller and larger than S1, respectively.

The most notable observation from Figs. 6–8 is that the SADRP activity for S1 to S3 was mainly confined within the putamen, whereas the RPE activity was mainly distributed within the caudate nucleus and ventral striatum. These nearly separate distributions of SADRP and RPE activities remained robustly consistent regardless of the differences in task difficulty from S1 to S3. Second, the number of voxels correlated with SADRP and RPE strongly depended on task difficulty in exactly the opposite manner: SADRP activity tended to be more prominent in the less stochastic task (S1) than in the more stochastic tasks (S2 and S3), whereas RPE activity both in the caudate nucleus and ventral striatum tended to exhibit stronger correlations in the more stochastic tasks (S2 and S3). More specifically, the number of voxels correlated with SADRP in S1, S2, and S3 was 683, 87, and 101, respectively, and the number correlated with RPE was 399, 864, and 565. Only the SADRP activity for S1 significantly overlapped RPE activity (SADRP had only 5 overlapping voxels with RPE for both S2 and S3). The number of overlapping voxels of SADRP for S1 with RPE for S1, S2, and S3 was 40, 180, and 107, respectively.

It is also important to know whether there is a common activation for SADRP and RPE across different task difficulties. To address this, we conducted a conjunction analysis (see METHODS) of SADRP and RPE over three sessions (S1–S3). Figure 9 overlays the results on a normalized brain image, where voxels correlated with SADRP in all sessions are shown in pink, whereas voxels correlated with RPE in all sessions are in green. The SADRP correlation was confined to the putamen in the vicinity of the anterior commissure. In contrast, the RPE correlation was localized in the caudate nucleus, again in the vicinity of the anterior commissure. Importantly, there was no overlap of correlation between SADRP and RPE.

To pinpoint the anatomical localizations of the correlation with SADRP and RPE and look into temporal characteristics of these brain activities, we examined a single subject's data and BOLD signal time-courses in early and late learning trials. Figure 10*A* shows SADRP correlation (red) and RPE correlation (cyan) while a typical subject was engaged in S2, which are overlaid on the subject's normalized structural image. Event-related BOLD signals averaged over the first and last 24 trials are also plotted at the peaks within the putamen (Fig. 10*B*) and the caudate nucleus (Fig. 10*C*). Consistent with the conjunction analysis, the individual subject analysis also shows that the correlations with SADRP and RPE were confined to the putamen and caudate nucleus, respectively. Event-related plots show that the BOLD signal in the putamen increased at fractal stimulus onset as learning proceeded, whereas the BOLD signal in the caudate nucleus decreased at the reward feedback timing. Thus this spatiotemporal feature of the BOLD signal is consistent with our hypothesis that brain activities in the putamen and caudate nucleus are mainly driven by SADRP and RPE, respectively.

Finally, the validity of the model-based correlation analysis depends on whether the activity in voxels indeed reflects the changes in SADRP or RPE or whether it comes from some other variables that are in turn correlated with either of these variables. To verify the reliability of this analysis, we carried out two additional multivariate regression analyses: one with both SADRP and AR, which basically form an increasing function, and the other with both RPE and change in SADRP (CSADRP), which basically form a decreasing function. Correlation with AR was found in the insula and inferior temporal cortex, whereas correlation with CSADRP was found only in the medial prefrontal cortices ($P < 0.001$; uncorrected for multiple comparisons). The correlation in the striatum with SADRP and RPE did not change by this inclusion of AR and CSADRP. Therefore within the context of this study, SADRP and RPE are more representative of activities in the putamen and caudate nucleus than AR and CSADRP, respectively.

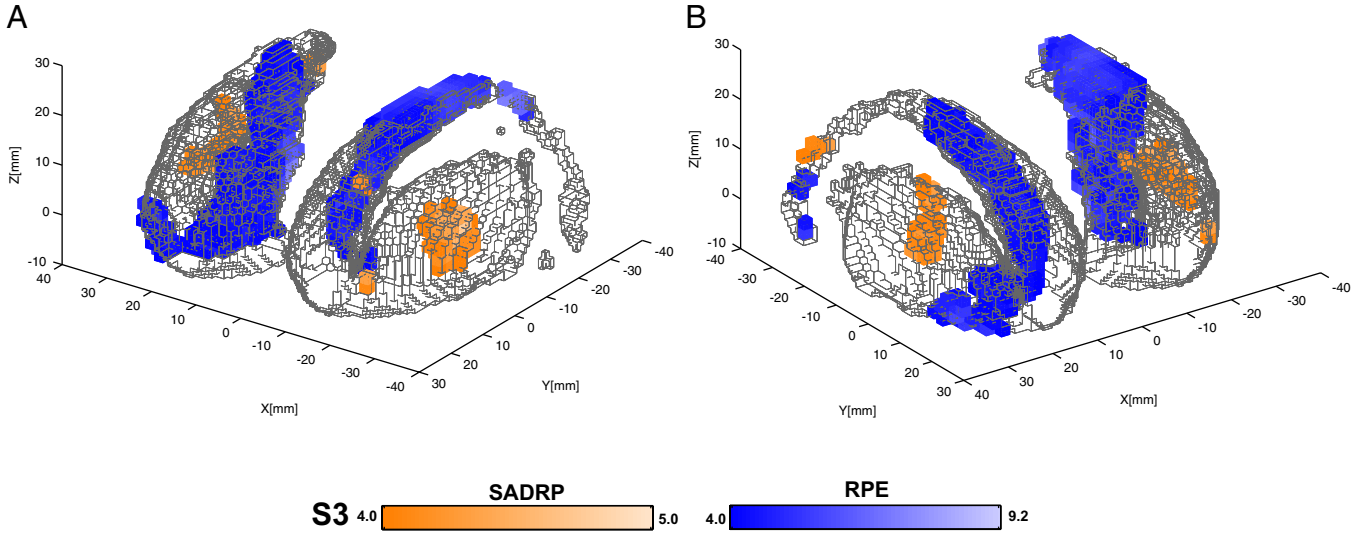

FIG. 8. Activity in the dorsal striatum correlated with SADRP (orange) and RPE (magenta) for S3. Figure is in the same format as Fig. 6. Peak T-values for SADRP activity in the left and right putamen were 4.95 and 4.50, for RPE activity in the left and right caudate nucleus were 9.19 and 7.40, and for the left and right ventral striatum activity were 4.95 and 5.15, respectively.

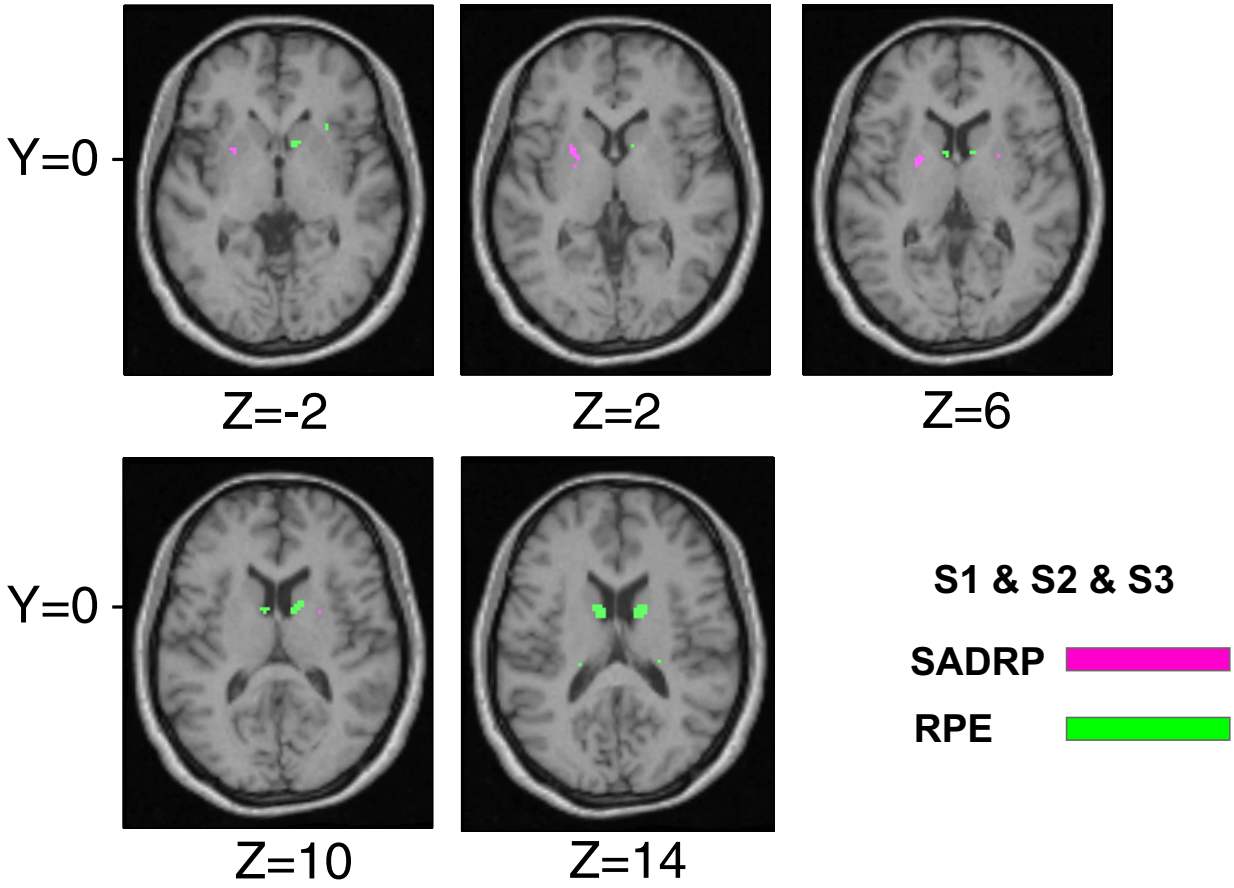

FIG. 9. Conjunction of activations for SADRP (pink) and RPE (green) across S1–S3 overlaid on the normalized structural image of a subject. *Y* and *Z* represent the MNI coordinates of the activations.

Other brain areas

Although the specific focus of this study was the striatum because numerous previous studies had suggested its central role in learning stimulus-action-reward associations, the activities of other brain regions were also found in the event-related correlations with SADRP and RPE (statistical threshold was set at $P < 0.001$, uncorrected for multiple comparisons, with the additional constraint that at least 10 contiguous voxels be included). Consistent correlations with SADRP were found for S1–3 in the bilateral superior parietal, dorsolateral prefrontal, dorsal premotor and occipital cortices, insula, thalamus, cerebellum, anterior cingulate cortex, supplementary motor area, and right superior temporal sulcus, whereas consistent RPE correlations for S1–3 were found in the bilateral superior parietal and occipital cortices, insula, hippocampus, anterior cingulate cortex, and right orbitofrontal, dorsolateral prefrontal, and dorsal premotor cortices. Unlike the putamen and caudate nucleus, none of the other brain regions correlated with both SADRP and RPE (i.e., the superior parietal, dorsolateral prefrontal, dorsal premotor and anterior cingulate cortices, and insula) exhibited any systematic differences in spatial activation pattern between SADRP and RPE, as seen by the nearly separate distributions throughout S1–3.

DISCUSSION

We conducted an event-related fMRI study with monetary reward to investigate the involvement of the putamen and caudate nucleus during stimulus-action-reward association learning. The results showed that activity in the putamen is mainly correlated with SADRP at the cue presentation, whereas activity in the caudate nucleus is mostly correlated with RPE at the reward feedback. This difference in the distribution of the correlations in the dorsal striatum suggests that the putamen acquires stimulus-action-dependent reward prediction dominantly, while the caudate nucleus, as well as the ventral striatum, is mainly engaged in the learning process controlled by comparing actual and predicted rewards. Although our previous work (Haruno et al. 2004) also dealt with fMRI examination of modular structures in the brain related to stochastic decision tasks, the significant new findings on the functional difference in the dorsal striatum, as well as the event-related task design and computational model–based analysis, are entirely new features of this study.

SADRP is critical for selecting an optimal behavior because an action as well as a contextual stimulus should be considered in predicting the amount of reward. The relevance of SADRP as the subject's internal representation in this study was indicated by the following observations. First, as shown in Figs. 3 and 4, the learning process simulated by the model based on SADRP (and consequently RPE) coincided with each subject's learning behavior. Second, the nearly mirror-image relationships between Fig. 5, *B* and *D*, indicate that SADRP explains behavioral learning better than does RPE. Third, the RPE calculated from the SADRP reflects the well-established finding that activity in the ventral striatum is strongly correlated

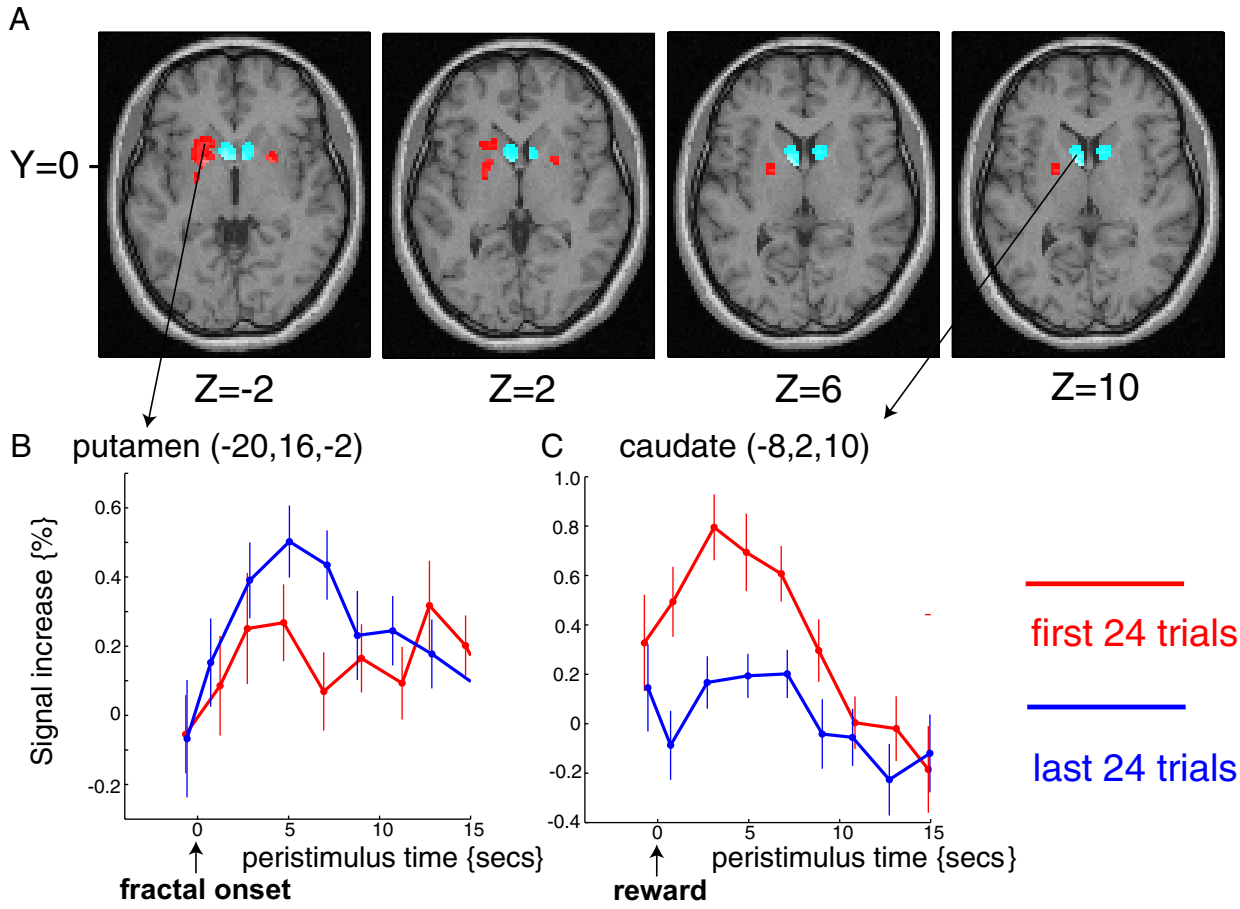

FIG. 10. Anatomical localization and time-course of brain activity for SADRP and RPE. *A*: correlated brain activity of a typical subject with SADRP (red) and RPE (cyan) for S2. Color bar is the same as in Fig. 7. *B*: event-related plot of the peak voxel in the putamen (−20,16,−2) at fractal onset. *C*: event-related plot of the peak voxel in the caudate nucleus (−8,2,10) at reward feedback. Red and blue lines in plots represent BOLD signals averaged over the 1st and last 24 trials, respectively.

with the TD error (Berns et al. 2001; Breiter et al. 2001; McClure et al. 2003; O'Doherty et al. 2003; Pagnoni et al. 2002). These observations together, with the fact that SADRP-correlated activity was bilateral (all subjects pushed a button with their right hand) and that brain activity purely related to the button-push was eliminated by subtracting the CONTROL activity, suggest that the SADRP correlated activity in the putamen represents the learning of stimulus-action-reward associations.

These correlations between putamen activity and SADRP and between caudate nucleus activity and RPE are consistent with their respective anatomical connections with the cortex: the anterior-intermediate putamen receives projections from the sensorimotor cortices, including the dorsal and ventral premotor cortices, the supplementary motor area, and the primary motor cortex (Alexander et al. 1990; Gerardin et al. 2003; Parthsarathy et al. 1992; Selemon and Goldman-Rakic 1985; Takada et al. 1998), whereas the caudate nucleus receives its inputs from frontal association areas, such as the dorsolateral prefrontal, orbitofrontal, and cingulate cortices (Alexander et al. 1990). Thus the medial-intermediate and anterior-intermediate putamen in the vicinity of the anterior commissure, which exhibit peak correlation with SADRP, are suitable locations for encoding stimulus-action-reward associations. This assumption is also supported by the fact that these are not only reward-related areas (Cromwell and Schultz 2003) but also motor-related areas, as a result of the projections from both the premotor cortex and the supplementary motor area. This general area of the putamen might be related to the integration of information on the expectation of reward with processes that mediate the actions leading to the reward. Similarly, the anatomical connections of the caudate nucleus suggest that it is appropriately located for dealing with the RPE. There are also some indications that reward and penalty are encoded by different neural substrates (Daw et al. 2002). Therefore we carried out separate analyses for positive and negative rewards and found that activity in the amygdala and hippocampus was correlated with the negative reward prediction error. In contrast, the brain regions activated for positive rewards were the same as those indicated by the current unified analysis, and the statistical significance became slightly weaker.

The view that the anterior-intermediate putamen acquires the stimulus-action-reward association is compatible with the results of recent electrophysiological studies with monkeys and the results of human imaging studies. After the completion of learning, a higher percentage of tonically active neurons (TANs) in the putamen respond to "go" signals for an action than in the caudate nucleus, especially when a reward is expected from the action (Yamada et al. 2004). Similarly, more prevalent activations preceding the trigger stimulus for an

action were found in projection neurons of the putamen (Cromwell and Schultz 2003). In the context of sequential motor learning, the posterior putamen was found to be more active when a monkey was conducting an already-learned motor sequence (Hikosaka et al. 1999, 2002; Miyachi et al. 1997, 2002) than when learning a new sequence. Similarly, a human PET study of sequential finger movement learning reported that the posterior putamen was activated when the sequential movements were well learned, whereas the intermediate putamen and caudate nucleus were activated during intermediate learning and new learning, respectively (Jeuptner and Weiller 1998; Jeuptner et al. 1997). Although, because of its limited temporal resolution, the PET study could not be focused on the timing of stimulus presentation, it is possible that the increase in the PET signal in the intermediate putamen represents the stimulus-action-reward association. The characteristics of fMRI that mainly reflect averaged synaptic inputs (Logothetis et al. 2001) (from motor-related areas in this experiment) may explain why our study highlighted the role of the putamen in stimulus-action-reward association more than previous electrophysiological studies. To identify a detailed computational mechanism executed in the putamen and caudate nucleus, it is essential to determine whether the dopamine system as well as the thalamostriatal loop (Smith et al. 2004) acts on these two structures equally or differently by conducting a PET (Zald et al. 2004) or electrophysiological study during stimulus-action-reward association learning.

This study focused on the difference between the putamen and caudate nucleus, and it is consistent with previous imaging studies based on TD models (Berns et al. 2001; McClure et al. 2003; O'Doherty et al. 2003; Pagnoni et al. 2002). Correlation with TD error was reported in the caudate nucleus in an instrumental conditioning task in addition to the ventral striatum, which was also activated in a classical conditioning task. This study revealed the correlation of activity with RPE in both the caudate nucleus and ventral striatum during stimulus-action-reward association learning (instrumental conditioning). In comparison with these studies, the main contribution of this study was to show the different involvement of the putamen and caudate nucleus during stimulus-action-reward association learning. Our results did show that a small number of voxels (5.8% of total correlated with SADRP only in S1) were correlated with both SADRP and RPE. These voxels exhibited BOLD signal time-courses that are analogous to the dopamine neurons of Schultz. That is, at the beginning of learning, the BOLD signal increase was marked at the timing of reward delivery, while in the later phase of learning, the BOLD signal increase was large at the visual stimulus timing and also remained at the reward delivery timing with a smaller amplitude. Thus one can argue that this small number of voxels exhibit similar time courses as the "TD error" encoded by dopamine neurons. However, we also emphasize that the majority of activated voxels were correlated with either SADRP at the timing of visual stimulus or RPE at the timing of reward delivery. This might be attributed to the fact that our task does not contain the feature of "temporal credit assignment," or the fMRI paradigm may not provide a high enough spatiotemporal resolution to examine this issue fully.

Although this study focused specifically on the contribution of the dorsal striatum during stimulus-action-reward association learning, other brain regions were also activated (see RESULTS). These regions activated by SADRP were consistent with the regions identified in previous human and monkey studies, i.e., the anterior cingulate cortex (Williams et al. 2004), prefrontal cortex (Barraclough et al. 2004; Matsumoto et al. 2003), and parietal cortex (Sugrue et al. 2004), suggesting that the dorsal striatum is a part of a large brain network involved in stimulus-action-reward association learning and subsequent decision making.

ACKNOWLEDGMENTS

We thank W. Schultz, K. Nakamura, M. Kimura and K. Toyama for helpful discussions and comments and T. Yoshioka and S. Tada for technical assistance.

GRANTS

This study was supported by NICT and by grants to M. Kawato from the Human Frontier Science Program. M. Haruno was partly supported by the Daiwa Anglo-Japanese Foundation.

REFERENCES

Alexander GE, Crutcher MD, and Delong MR. Basal ganglia thalamocortical circuits: parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. *Prog Brain Res* 85: 119–146, 1990.

Barraclough DJ, Conroy ML, and Lee D. Prefrontal cortex and decision making in a mixed-strategy game. *Nat Neurosci* 7: 404–410, 2004.

Barto AG, Sutton RS, and Anderson CW. Neuron-like elements that can solve difficult learning control problems. *IEEE Trans Syst Man Cybern* 13: 835–846, 1983.

Berns GS, McClure MS, Pagnoni G, and Montague PR. Predictability modulates human brain response to reward. *J Neurosci* 21: 2793–2798, 2001.

Bertsekas DP and Tsitsiklis JN. *Neuro-Dynamic Programming.* Belmont, MA: Athena Scientific, 1996.

Breiter HC, Aharon I, Kahneman D, Dale A, and Shizgal P. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* 30: 619–639, 2001.

Brown J, Bullock D, and Grossberg S. How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *J Neurosci* 19: 10502–10511, 1999.

Cromwell HC and Schultz W. Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *J Neurophysiol* 89: 2823–2838, 2003.

Daw ND, Kakade S, and Dayan P. Opponent interactions between serotonin and dopamine. *Neural Netw* 15: 603–616, 2002.

Dayan P, Kakade S, and Montague RP. Learning and selective attention. *Nat Neurosci* 3: 1218–1223, 2000.

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

Friston KJ, Holmes AP, Worsley K, Poline JB, Frith C, and Frackowiak RSJ. Statistical parametric maps in functional brain imaging: a general linear approach. *Hum Brain Map* 2: 189–210, 1995.

Gerardin E, Lehericy S, Pochon JB, Tezenas du Montcel S, Mangin JF, Poupon F, Agid Y, Le Bihan D, and Marsault C. Foot, hand, face and eye representation in the human striatum. *Cereb Cortex* 13: 162–169, 2003.

Haruno M, Kuroda T, Doya K, Toyama K, Kimura M, Samejima K, Imamizu H, and Kawato M. A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task. *J Neurosci* 24: 1660–1665, 2004.

Hikosaka O, Nakahara H, Rand MK, Sakai K, Lu X, Nakamura K, Miyachi S, and Doya K. Parallel neural networks for learning sequential procedures. *Trends Neurosci* 22: 464–471, 1999.

Hikosaka O, Nakamura K, Sakai K, and Nakahara H. Central mechanisms of motor skill learning. *Curr Opin Neurobiol* 12: 217–222, 2002.

Hollerman JR and Schultz W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat Neurosci* 4: 304–309, 1998.

Houk JC, Adams JL, and Barto AG. A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: *Models of Information Processing in the Basal Ganglia*, edited by Houk JC, Davis JL, and Beiser DG. Cambridge, MA: MIT Press, 1995, p. 249–270.

Jeuptner M, Frith CD, Brooks DJ, Frackowiak RSJ, and Passingham RE. Anatomy of motor learning. II. Subcortical structures and learning by trial and error. *J Neurophysiol* 77: 1325–1337, 1997.

Jeuptner M and Weiller C. A review of differences between basal ganglia and cerebellar control of movements as revealed by functional imaging studies. *Brain* 121: 1437–1449, 1998.

Kawagoe R, Takikawa Y, and Hikosaka O. Expectation of reward modulates cognitive signals in the basal ganglia. *Nat Neurosci* 1: 411–416, 2001.

Knutson B, Adams MC, Fong WG, and Hommer D. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci* 159: 1–5, 2001.

Logothetis NK, Pauls J, Augath M, Trinath T, and Oeltermann A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412: 128–130, 2001.

Matsumoto N, Hanakawa T, Maki S, Graybiel AM, and Kimura M. Role of nigrostriatal dopamine system in learning to perform sequential motor tasks in a predictive manner. *J Neurophysiol* 82: 978–998, 1999.

Matsumoto K, Suzuki W, and Tanaka K. Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* 30: 229–232, 2003.

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

Miyachi S, Hikosaka O, Miyashita K, Karadi Z, and Rand MK. Differential roles of monkey striatum in learning of sequential hand movement. *Exp Brain Res* 115: 1–5, 1997.

Miyachi S, Hikosaka O, and Lu X. Differential activation of monkey striatal neurons in the early and late stages of procedural learning. *Exp Brain Res* 146: 122–126, 2002.

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

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

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

Pagnoni G, Zink CF, Montague PR, and Berns GS. Activity in human ventral striatum locked to errors of reward prediction. *Nat Neurosci* 5: 97–98, 2002.

Parthsarathy HB, Schall JD, and Graybiel AM. Distributed but convergent ordering of corticostriatal projections: analysis of the frontal eye field and the supplementary eye field in the macaque monkey. *J Neurosci* 12: 4468–4488, 1992.

Schultz W, Apicella P, Scarnati E, and Ljungberg T. Neuronal activity in monkey ventral striatum related to the expectation of reward. *J Neurosci* 12: 4595–4610, 1992.

Schultz W and Dickinson A. Neuronal coding of prediction errors. *Annu Rev Neurosci* 23: 473–500, 2000.

Schultz W, Tremblay L, and Hollerman JR. Changes in behavior-related neuronal activity in the striatum during learning. *Trends Neurosci* 26: 321–328, 2003.

Selemon LD and Goldman-Rakic PS. Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. *J Neurosci* 5: 776–794, 1985.

Shidara M, Aigner TG, and Richmond BI. Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J Neurosci* 18: 2613–2625, 1998.

Smith Y, Raju DV, Pare JF, and Sidibe M. The thalamostriatal system: a highly specific network of the basal ganglia circuitry. *Trends Neurosci* 27: 520–527, 2004.

Sugrue LP, Corrado GS, and Newsome WT. Matching behavior and the representation of value in the parietal cortex. *Science* 304: 1782–1787, 2004.

Sutton RS and Barto AG. *Reinforcement Learning.* Cambridge, MA: MIT Press, 1998.

Takada M, Tokuno H, Nambu A, and Inase M. Corticostriatal projections from the somatic motor areas of the frontal cortex in the macaque monkey: segregation versus overlap of input zones from the primary motor cortex, the supplementary motor area, and the premotor cortex. *Exp Brain Res* 120: 114–128, 1998.

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

Talairach J and Tournoux P. *Co-Planar Stereotaxic Atlas of the Human Brain*. New York: Thieme, 1998.

Tremblay I, Hollerman JR, and Schultz W. Modifications of reward expectation-related neuronal activity during learning in primate striatum. *J Neurophysiol* 80: 964–977, 1998.

Tricomi EM, Delgado MR, and Fiez JA. Modulation of caudate activity by action contingency. *Neuron* 41: 281–292, 2004.

Williams ZM, Bush G, Rauch SL, Cosgrove GR, and Eskandar EN. Human anterior cingulate neurons and the integration of monetary reward with motor responses. *Nat Neurosci* 12: 1376–1380, 2004.

Yamada H, Matsumoto N, and Kimura M. Tonically active neurons in the primate caudate nucleus and putamen differently encode instructed motivational outcomes of action. *J Neurosci* 24: 3500–3510, 2004.

Young P. *Recursive Estimation and Time Series.* New York: Springer-Verlag, 1984.

Zald DH, Boileau I, El-Deared W, Gunn R, McGlone F, Dichter GS, and Dagher A. Dopamine transmission in the human striatum during monetary reward tasks. *J Neurosci* 24: 4105–4112, 2004.