History- and Current Instruction-Based Coding of Forthcoming Behavioral Outcomes in the Striatum

Hiroshi Yamada, Naoyuki Matsumoto, and Minoru Kimura
Department of Physiology, Kyoto Prefectural University of Medicine, Kyoto, Japan

Submitted 11 July 2007; accepted in final form 8 October 2007

Yamada H, Matsumoto N, Kimura M. History- and current instruction-based coding of forthcoming behavioral outcomes in the striatum. J Neurophysiol 98: 3557–3567, 2007. First published October 10, 2007; doi:10.1152/jn.00779.2007. Animals optimize behaviors by predicting future critical events based on histories of actions and their outcomes. When behavioral outcomes like reward and aversion are signaled by current external cues, actions are directed to acquire the reward and avoid the aversion. The basal ganglia are thought to be the brain locus for reward-based adaptive action planning and learning. To understand the role of striatum in coding outcomes of forthcoming behavioral responses, we addressed two specific questions. First, how are the histories of reward and aversion used for encoding forthcoming outcomes in the striatum during a series of instructed behavioral responses? Second, how are the behavioral responses and their instructed outcomes represented in the striatum? We recorded discharges of 163 presumed projection neurons in the striatum while monkeys performed a visually instructed lever-release task for reward, aversion, and sound outcomes, whose occurrences could be estimated by their histories. Before outcome instruction, discharge rates of a subset of neurons activated in this epoch showed positive or negative regression slopes with reward history (24/44), that is, to the number of trials since the last reward trial, which changed in parallel with reward probability of current trials. The history effect was also observed for the aversion outcome but in far fewer neurons (3/44). Once outcomes were instructed in the same task, neurons selectively encoded the outcomes before and after behavioral responses (reward, 46/70; aversion, 6/70; sound, 6/70). The history- and current instruction-based coding of forthcoming behavioral outcomes in the striatum might underlie outcome-oriented behavioral modulation.

INTRODUCTION

Optimizing behavior to achieve a specific goal by predicting future critical events is a hallmark of human and animal intelligence. In a new environment where routes to a distant goal are unknown, a trial-and-error approach is a common strategy for seeking rewards and avoiding aversions (Thorndike 1898). Based on rewards and aversions experienced after past actions, the outcomes of present actions can be predicted. Through experiencing a large number of actions, their associated sensory events, and their outcomes, we learn the linkages among them.

Prospective coding of future events, such as the coding of visual objects associated with another currently present object, occurs in the cerebral cortex (Rainer et al. 1999; Sakai and Miyashita 1991), thalamus (Komura et al. 2001), and basal ganglia (Lauwereyns et al. 2003). Neuronal activity during the exploration and exploitation of spatial locations for reward occurs in the striatum (Barnes et al. 2005). Theories of reinforcement learning have proposed algorithms in which organisms estimate the reward that actions might yield and choose actions by comparing the reward values of options based on the histories of particular actions and rewards (Sutton and Barto 1998). Reinforcement learning models of the basal ganglia (Doya 2000; Houk et al. 1995; O’Doherty et al. 2004) are supported by the observations that dopamine neurons encode reward-prediction errors (Bayer and Glimcher 2005; Fiorillo et al. 2003; Morris et al. 2004; Nakahara et al. 2004; Satoh et al. 2003), and that striatal neuron activity represents reward values of external stimuli (Cromwell and Schultz 2003; Kawagoe et al. 1998) as well as reward values that actions might yield (Samejima et al. 2005). Thus it is suggested that the dorsal striatum may be the brain locus for history-based coding of forthcoming action outcomes, especially reward. On the other hand, it is also often necessary to predict and avoid aversion outcomes. Whether, and if so, the striatum is involved in encoding aversion outcomes remains unclear, although striatal neurons do respond to aversive stimuli (Blazquez et al. 2002; Ravel et al. 2003; Yamada et al. 2004).

The present study addressed two specific questions to understand the roles of the striatum in coding outcomes of forthcoming behavioral responses. First, how are the histories of reward and aversion used for encoding forthcoming outcomes in the striatum during a series of instructed behavioral responses? Second, how are the behavioral responses and their instructed outcomes represented in the striatum? We examined the activity of presumed projection neurons in the striatum while two Japanese monkeys performed a visually guided lever-release task for reward, aversion, and sound outcomes, the occurrences of which could be estimated by their histories in the preceding several trials. Our results support the view that striatal projection neurons encode outcomes of forthcoming behavioral responses, especially reward, based on their histories, and modulated their behavioral response-related activity in an instructed outcome-selective manner.

METHODS

Experimental animals

We used two Japanese monkeys (Macaca fuscata; monkey DA, male, 5.6 kg and monkey AI, female, 6.0 kg). All animal procedures were approved by the Animal Care and Use Committee of Kyoto Prefectural University of Medicine and conformed to National Institutes of Health guidelines.

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.
Behavioral task

The monkeys sat on a primate chair facing a wooden panel and performed a visually guided lever-release task for the following outcomes: a water reward, avoiding an airpuff on their faces, or hearing a beep sound (Fig. 1A; see also Yamada et al. 2004). They depressed a hold lever at their own pace and were required to fixate their eyes on a fixation point (FP) [a small light-emitting diode (LED) in the center of the panel]. Another larger LED, as an instruction stimulus of response outcome, was turned on to the right of the fixation LED with two different colors presented sequentially (15 and 5° apart in monkey DA and AI, respectively). The location of the large LED was closer to the FP for monkey AI than for monkey DA because it was difficult for monkey AI to maintain her gaze on the FP with the LED at a greater distance from the FP. First, a green LED cued monkeys to wait for a subsequent instruction with the hold lever depressed (wait cue). After 1.2–1.8 s, the green LED changed color and instructed the monkeys of the outcome of the trial (Outcome Instruction). When the outcome instruction was turned off (GO), the monkeys released the hold lever quickly (reaction time <600 ms). When the outcome instruction was red, a drop of reward water (0.2 ml) was delivered into the mouth about 600 ms after the lever release. When it was yellow, a beep sound (1,500 Hz, 200 ms, 70 psi, 30 ms) only when the lever-release responses were delayed (>600 ms). When it was blue, an airpuff was delivered to the right side of the face (26 psi, 30 ms) only when the lever-release responses were delayed (>600 ms). Therefore, based on the experience of the trial, monkeys repeated the trial with the same outcome. The number of errors (fixation break, lever release before GO signal, or delayed lever release) were considered to be fixation errors. If a trial was aborted due to errors (fixation break, lever release before GO signal, or delayed lever release), monkeys repeated the trial with the same outcome. The outcome instruction was either preceded (monkey DA) or followed by the wait cue (monkey AI) (Yamada et al. 2004).

Each of the three outcomes occurred twice, in random order, within a six-trial block (Fig. 1B). Therefore based on the experience of the last several trials, monkeys were able to roughly estimate the probabilities of reward, aversion, and sound trials before the outcome instructions actually appeared. The maximum possible number of consecutive unrewarded trials (aversion and sound) was eight, when two reward trials appeared at the beginning and end of consecutive two six-trial blocks. Thus the number of trials since the last reward trial [Trials since Last Reward (TLR)] varied from one to nine. Similarly, the number of trials since the last aversion trial [Trials since Last Aversion (TLA)] and the number of trials since the last sound trial [Trials since Last Sound (TLS)] also varied from one to nine. Only error-free trials were used to calculate the TLR, TLA, and TLS because the same outcome was repeated after errors.

Recording

We recorded single-neuron discharges from the caudate nucleus and putamen in the left hemisphere of two monkeys (the region from A13 to A22 of the atlas of Japanese monkey brain in the Horsley-Clark coordinate system). We used epoxy-coated tungsten microelectrodes to record and a template-matching algorithm to isolate single-neuron discharges. Striatal presumed projection neurons and tonically active neurons (TANs, presumed cholinergic interneurons) were identified based on their background discharge rates and action potential waveforms (Aosaki et al. 1995). Striatal projection neurons showed low spontaneous firing rates (<2 spikes/s) and phasic discharges in relation to one or more behavioral task events (Kimura et al. 1996). Although we think that most of the “presumed” projection neurons are GABAergic projection neurons, it was still possible that the population of neurons contained very small number of γ-aminobutyric acid (GABA) interneurons. Discharges of single neurons were recorded during about 75 trials (25 trials for each outcome). Eye movements were monitored by measuring the corneal reflections of an infrared light beam by a video camera with a time resolution of 4 ms. Recordings started when the monkeys had performed the behavioral task at correct performance rates >80% in each outcome trial. This required 1 mo for monkey DA and 3 mo for monkey AI after the aversion trials were last introduced, following reward and sound outcomes.

Analysis of behavioral measures for task performance

Neuronal and behavioral data were collected during 44,909 trials in monkey DA and 8,887 trials in monkey AI. To examine effects of outcome instructions on task performance, correct performance rates, reaction times for lever-release responses, and number of error trials were compared among the three outcome conditions using multiple two-sample comparisons corrected by Bonferroni method to control the family-wise significance level. To examine effects of the histories of reward, aversion, and sound on task performance, task start times (latency from release of hold lever in the preceding trial) to depression of the lever to start the next trial) during all recording sessions, and rate of errors before the outcome instructions in each single recording session were measured. Trials with unusually long task start times (>3 s) were not included (0.3% in monkey DA and 4% in monkey AI). Almost all of the errors were due to fixation breaks before the outcome instructions. Effects of the TLR and TLA on the behavioral measures were examined using the following regression model

$$Y = b_0 + b_{TLR} \cdot TLR + b_{TLA} \cdot TLA + \epsilon$$  \hspace{1cm} (1)

where $b_0$ and error are intercept and residual, respectively. Statistical significance of the regression coefficient was evaluated using t-test at $P < 0.05$ (whether the probabilities of $b_{TLR}$ and $b_{TLA}$ were zero). Because the histories of three outcomes were not independent (see Correlation between variables and linear regressions), effects of the TLS, as a control, on the behavioral measures were also examined instead of TLA using the following regression model

$$Y = b_0 + b_{TLR} \cdot TLR + b_{TLS} \cdot TLS + \epsilon$$  \hspace{1cm} (2)

However, the coefficients to TLR derived from the two models were not significantly different (not shown) and thus we adopted the former model (Eq. 1).
Analysis of neuronal discharge rates

Histograms of neuronal discharge rates were constructed before and after particular behavioral events. Average discharge rates were determined separately during nine task epochs in each trial: for 300 ms following the hold-lever depression, for 500 ms preceding the wait cue, for 500 ms following the wait cue, for 500 ms preceding the outcome instructions, for 500 ms following the outcome instructions, for 500 ms preceding the lever release, for 300 ms following the lever release, for 300 ms preceding the outcome, and for 500 ms following the outcome. In each epoch, an increase in discharge rate was regarded as significant when it was higher than the background rate (Wilcoxon two-sample test, P < 0.05), defined in this study as the lowest average discharge rate across nine task epochs and >2.0 spikes/s.

To examine how behavioral responses and their forthcoming outcomes affected neuronal discharge rates, multiple regression analyses were applied to neurons that were active during at least one task epoch other than the outcome epochs. Neuronal discharge rates (F) were fitted by the following regression model including TLR and TLA, and current outcomes

\[ F = b_0 + b_{TLR} \text{TLR} + b_{TLA} \text{TLA} + b_{REW} \text{REW} + b_{AVE} \text{AVE} + \text{error} \]  (3)

where REW and AVE took a value of either 1 or 0 as the presence or absence of reward and aversion. If the TLR or TLA values were >5, which occurred in <10% of all trials, the trials were pooled as TLR ≥5 or TLA ≥5. If the regression coefficients \( b_{TLR} \) or \( b_{TLA} \) were not zero at \( P < 0.05 \), neuronal activity was regarded as being significantly modulated by the TLR or TLA. Effects of the TLS, as a control, on the neuronal discharges were also examined instead of TLA by

\[ F = b_0 + b_{TLR} \text{TLR} + b_{TLD} \text{TLD} + b_{REW} \text{REW} + b_{SOU} \text{SOU} + \text{error} \]  (4)

where REW and SOU took a value of either 1 or 0 as the presence or absence of reward and sound. The coefficients to TLR derived from the two models were not significantly different (not shown) and thus we adopted the former model (Eq. 3).

If \( b_{REW} \), \( b_{AVE} \), or \( b_{SOU} \) values were not zero at \( P < 0.05 \), the neuronal activity was regarded as depending on current outcomes. To define which types of current outcomes a particular neuron encoded (reward, aversion, sound, or both reward and aversion), the neuronal discharge rates in reward and aversion trials were compared with sound trials (two-sample t-test, \( P < 0.05 \)).

To examine the effect of the behavioral measures on the neuronal discharge rates, the following analyses were conducted. For neurons that were active during the task epochs before the outcome instructions, we used regression models that incorporated task start times (TSTs) into the above-mentioned models, for example

\[ F = b_0 + b_{TST} \text{TST} + b_{TLR} \text{TLR} + b_{TLA} \text{TLA} + b_{REW} \text{REW} + b_{AVE} \text{AVE} + \text{error} \]  (5)

For neurons that were active during the task epochs after the outcome instructions, we used regression models that included the reaction times from the GO signal (RTs), for example

\[ F = b_0 + b_{TST} \text{TST} + b_{TLR} \text{TLR} + b_{TLA} \text{TLA} + b_{REW} \text{REW} + b_{AVE} \text{AVE} + b_{RT} \text{RT} + \text{error} \]  (6)

The number of neurons modulated by TLR, TLA, and TLS were compared by Fisher’s exact probability test at \( P < 0.05 \).

Correlation between variables and linear regressions

TLR, TLA, and TLS were not independent, but were moderately correlated with each other (Table 1). Thus it was not necessarily appropriate to include all three variables for outcome histories in one formula (Grafen and Hails 2002). We first examined modulation of neuronal activity by TLR and TLA. Then, the effect of TLS as a control was examined using another model, as described earlier (Eqs. 2 and 4). Furthermore, the stepwise regression, which is one simple way of variable selection (Padoa-Schioppa and Assad 2006), was also used to compare the effects of each variable (TLR, TLA, REW, AVE, and SOU) on neuronal discharge rates. The forward stepwise regression was performed at \( P < 0.05 \). The numbers of neurons modulated by TLR, TLA, and TLS were not significantly different between above-mentioned models and the stepwise regression. We did not remove any regressors in previous analyses using Eqs. 1 to 6.

Histology

At the end of all recording experiments, small electrolytic lesions were made both in the caudate nucleus and in the putamen. Direct anodal current (20 μA) was passed for 30 s through tungsten micro-electrodes. The monkeys were deeply anesthetized with Nembutal (60 mg/kg, administered intraperitoneally) and were perfused transcardially with 10% formalin in 0.9% NaCl solution. Coronal sections of the striatum, 50 μm in thickness, were stained with cresyl violet. Electrode tracks through the striatum were reconstructed on the histology sections using the electrolytic lesion marks as reference points, and the recording sites of striatal projection neurons were identified.

Results

Influences of instruction of response outcomes on task performance

Monkeys learned to release a lever for three different outcomes (Fig. 1A; see also Yamada et al. 2004): a water reward, a beep sound, or avoiding an aversive airpuff on the face by quicker responses. The monkeys performed single behavioral responses (lever release) for three different outcomes. After several months of task performance, each of the three outcome instructions had significant influences on task performance (Table 2). In monkey DA (where the outcome instruction was preceded by the wait cue), reaction times for the lever-release response were shorter in reward and aversion trials than in sound trials, and those in reward trials were shorter than those in aversion trials (\( P < 0.01 \), Bonferroni correction). Rates of release errors before the Go signal were higher in aversion trials than in sound trials. On the other hand, monkey AI (where the wait cue followed the outcome instruction) performed the task with shorter reaction times than monkey DA, and the reaction times in aversion trials were shorter than those in sound trials. The behavioral responses were quicker after the aversion outcome was newly introduced in addition to the reward and sound (reaction times before introduction of aversion trial: 320.4 ± 1.83 ms, average of reward and sound trials ± SE). This indicated that monkey AI changed her

---

Table 1. Correlation matrix among histories of reward, aversion, and sound trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Reward</th>
<th>Aversion</th>
<th>Sound</th>
<th>Reward</th>
<th>Aversion</th>
<th>Sound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monkey DA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reward</td>
<td>1.000</td>
<td>-0.391</td>
<td>-0.391</td>
<td>1.000</td>
<td>-0.380</td>
<td>-0.385</td>
</tr>
<tr>
<td>Aversion</td>
<td>-0.391</td>
<td>1.000</td>
<td>-0.377</td>
<td>-0.380</td>
<td>1.000</td>
<td>-0.391</td>
</tr>
<tr>
<td>Sound</td>
<td>-0.391</td>
<td>-0.377</td>
<td>1.000</td>
<td>-0.385</td>
<td>-0.391</td>
<td>1.000</td>
</tr>
</tbody>
</table>

---

Values are correlation coefficients among histories of reward (Trials since Last Reward), aversion (Trials since Last Aversion), and sound (Trials since Last Sound) in monkey DA and Monkey AI during all recording sessions.
strategy to react to the GO signal as quickly as possible after the introduction of aversion trials. Rates of release errors before the GO signal were lower in reward trials than in sound trials. Because aversive face airpuff was applied to the monkeys only when the behavioral response was delayed, monkeys received an airpuff only a few times a day after learning (monkey DA; 2.0%, monkey AI; 0.52%). The observed behavioral measures of task performance under the three outcomes indicated that the instruction served as a predictor of outcomes, although the two monkeys did not perform in the same way.

Prediction of reward and aversion outcomes based on their histories

Each of the three instructions associated with their respective outcomes occurred twice within a six-consecutive-trial block (Fig. 1B). Therefore the probability of reward increased monotonically with the increase in the TLR (Fig. 1D). On the other hand, occurrence of large TLR decreased compared with small TLR because it was reset by individual reward trials in less than about six trials (Fig. 1C). Likewise, the probability of aversion trials increased monotonically with the increase in the TLA. The monkeys’ task start times became shorter as the TLR increased (Fig. 2, A and B, Table 3). In addition, the number of errors before the instruction of outcomes decreased as the TLR increased (Fig. 2, C and D), whereas the task start times or number of errors increased selectively to the trials with TLR = 1 (Fig. 2, A and D). Thus the reward history (TLR) significantly modulated these behavioral measures of task performance in both of the monkeys. In contrast, the aversion history (TLA) had only small influences on these behavioral measures (Table 3), and the influence of the TLA was not significantly different from that of the TLS (comparision of regression slope, P > 0.05 in all cases).

Neuronal coding before instruction of current outcomes reflects past trial information since last reward

We recorded spike discharges from 163 presumed projection neurons in the caudate nucleus and putamen. In 82 of them, the discharge rates increased during one or a few task epochs in each trial: following the hold-lever depression, preceding and following the task cue, preceding and following the outcome instructions, and preceding and following the lever release. We examined, first, whether and how the TLR and TLA influenced striatal neuron discharges. Figure 3A shows activity of a neuron in the caudate nucleus that exhibited gradually increasing discharges up to the outcome instruction and rapid decline after the instruction. The magnitude of the discharge rate had a positive regression slope with TLR (Fig. 3B, multiple regression analysis: slope = 3.02, P = 0.032, R^2 = 0.116), but not with TLA (slope = −1.14, P = 0.43, Fig. 3C). Thus the increase in discharge rates across trials (TLR) seemed to occur in parallel with the increase in reward probability (Fig. 3B). Indeed, the neuronal discharge rates had significant positive regression slope with expected reward probability calculated from the data throughout the neuronal recording period in Fig. 1D (slope = 36.6, P = 0.021, R^2 = 0.137). Reward history also influenced the discharge rates during depression of the hold lever, which occurred at the beginning of each trial. A putamen neuron in Fig. 3D exhibited a phasic increase in discharge rate just after the lever depression to start trials. The discharge rate had a positive regression slope with TLR (Fig. 3E, slope = 5.25, P < 0.001, R^2 = 0.313), whereas the regression slope with TLA was not significant (Fig. 3F, slope = −0.83, P = 0.529). In 82 neurons that discharged during 182 task epochs, we examined the influence of TLR and TLA on discharge rates using multiple regression analyses.

TABLE 2. Influence of current outcomes on behavioral measures

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Monkey DA</th>
<th>Monkey AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction time to GO signal, ms</td>
<td>Reward</td>
<td>266.7 (0.56)*†</td>
</tr>
<tr>
<td></td>
<td>Aversion</td>
<td>292.1 (0.71)§</td>
</tr>
<tr>
<td></td>
<td>Sound</td>
<td>295.5 (0.71)</td>
</tr>
<tr>
<td>Release error before GO signal, %</td>
<td>Reward</td>
<td>1.79 (0.20)†</td>
</tr>
<tr>
<td></td>
<td>Aversion</td>
<td>3.70 (0.34)§</td>
</tr>
<tr>
<td></td>
<td>Sound</td>
<td>2.31 (0.20)</td>
</tr>
</tbody>
</table>

Values are means ± SE. *: significant difference between reward and sound trials at P < 0.01 (Bonferroni correction); §: significant difference between aversion and sound trials at P < 0.01; †: significant difference between reward and aversion trials at P < 0.01.

TABLE 3. Modulation of behavioral measures by reward, aversion, and sound histories

<table>
<thead>
<tr>
<th>History</th>
<th>Monkey DA</th>
<th>Monkey AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task start time</td>
<td>Reward</td>
<td>−0.210**</td>
</tr>
<tr>
<td></td>
<td>Aversion</td>
<td>0.030**</td>
</tr>
<tr>
<td></td>
<td>Sound</td>
<td>0.043**</td>
</tr>
<tr>
<td>Error rates before outcome instruction</td>
<td>Reward</td>
<td>−0.161**</td>
</tr>
<tr>
<td></td>
<td>Aversion</td>
<td>−0.077**</td>
</tr>
<tr>
<td></td>
<td>Sound</td>
<td>−0.089**</td>
</tr>
</tbody>
</table>

Values are standardized regression coefficients for histories of reward (Trials since Last Reward), aversion (Trials since Last Aversion), and sound (Trials since Last Sound) in relation to the behavioral measures (see METHODS). **: statistical significance at P < 0.001 (multiple regression analysis).
Supplemental Table 1. The number of neurons modulated by number of neurons (6 in monkey DA and 5 in monkey AI; whereas TLA influenced discharge rates in much smaller number of neurons (3/44). In the 24 of 44 neurons (17/32 in monkey DA and 7/12 in monkey AI), exact probability test; TLR was larger than that modulated by TLS (11/82, Fisher’s P from that by TLS (0.99). Thus the neuronal discharge rates of striatal neurons were predominantly modulated by reward history compared with aversion history.

During task epochs before the outcome instruction, discharge rates had significant positive or negative regression slopes with TLR in 24 of 44 neurons (17/32 in monkey DA and 7/12 in monkey AI, Fig. 4, A–C). In the 24 neurons, no neurons showed significant regression slope with TLA. Occurrence of reward history-dependent modulation in the Wait epoch before outcome instruction (Fig. 4A; 9/16 in monkey DA) tended to be higher than that in the Wait epoch after instruction (Fig. 4B; 3/16 in monkey AI, P = 0.066). In the 24 neurons modulated by TLR before outcome instruction, 15 and 9 neurons showed positive and negative regression slopes with TLR, respectively (Fig. 5, A and B). Thus the discharge rates of a subset of striatal neurons were modulated by reward history (TLR) before the outcome instructions in both monkeys. The activity modulated by reward history may reflect estimated probability of reward in current trials while the monkeys were predicting forthcoming outcomes before the instructions appeared. In contrast, aversion history (TLA) influenced discharge rates in much smaller number of neurons (3/44). About 35% (16/44) of neurons were never modulated by either one of histories of reward, aversion, and sound.

Coding of behavioral response and its outcomes after an explicit instruction of current outcomes

Once the outcomes of current trials were instructed, monkeys initiated behavioral responses by releasing the lever to acquire the water, avoid the airpuff, or hear the beep sound. The percentage of neurons showing a significant regression slope with TLR declined after the outcome instruction (3/39 neurons, P < 0.001 in monkey DA and 7/31, P = 0.035 in monkey AI) (11/82, Fisher’s exact probability test; P = 0.001), whereas the number of neurons modulated by TLA was not significantly different from that by TLS (P = 0.99). Thus the neuronal discharge rates of striatal neurons were predominantly modulated by reward history compared with aversion history.

1 The online version of this article contains supplemental data.
monkey AI; Fig. 4, A, B, and D). The percentage of neurons modulated by TLR after the outcome instruction was not significantly different from that of neurons modulated by TLS \((3/39 \text{ vs. } 6/39, P = 0.298 \text{ in monkey DA and } 7/31 \text{ vs. } 2/31, P = 0.147 \text{ in monkey AI})\). About 70% (48/70) of neurons were never modulated by either one of histories of reward, aversion, and sound. Thus we examined whether the striatal neurons encode instructed outcomes during behavioral performances after the explicit instruction of outcomes. Many neurons exhibited instructed outcome-selective activity, while exhibiting no clear modulation by TLR. For example, a putamen neuron showed burst discharges at the lever release response after the instruction of reward outcome (Fig. 6A, reward vs. sound, \(P < 0.001\)). Following the instructions of aversion and sound outcomes, however, the neuron exhibited almost no activation (Fig. 6, C and D, aversion vs. sound, \(P = 0.696\)). The discharge rates of this neuron during trials for reward did not show a significant regression slope with TLR (Fig. 6B, slope \(P = 0.111, P = 0.32\)). This is in sharp contrast to the activity surrounding the depression of the lever at the beginning of the trial, which reflects past information about TLR. We examined the modulation of discharge rates by types of outcome in current trials of 70 neurons that were active during task epochs after the outcome instruction. Significant modulation of discharge rates occurred in a large number of neurons by the outcomes of current behavioral responses, especially for the reward outcome (43/70, Fig. 8C). The reward outcome–selective neurons showed higher discharges in reward trials than in sound trials, with no significant difference between aversion and sound trials (24/43, Figs. 6 and 7B), although slightly more than half of the neurons (23/43) showed lower discharges in reward trials than in sound trials, with no significant difference between aversion and sound trials (Fig. 7, A and C). The

FIG. 5. Reward history–based modulations of neuronal discharge rate before the outcome instruction. Average discharge rate of each neuron plotted as a function of TLR. There were 2 groups of neurons exhibiting increases (A) and decreases (B) in discharge rates with increasing TLR. In one neuron, discharge rates during 2 task epochs before the outcome instruction are plotted separately in B because significant modulation of discharges by TLR occurred. Circles and triangles represent data from monkeys DA and AI, respectively. Black diamonds represent mean values.

FIG. 6. Lever release response–related activity was modulated by the instructed outcome of current trials. A: lever release–related increase of discharge rates of a putamen neuron in monkey DA occurred at all trials with different TLRs. Shown is activity during reward trials. B: plots of discharge rates during the hatched time window in A as a function of TLR. C and D: same neuron as in A and B, but for activity during the aversion and sound trials (pooled). Format of illustration is the same as that in Fig. 3, A and B.

FIG. 7. Neuronal discharges selectively modulated by each instructed outcome. A: activity of a putamen neuron in monkey AI showing a lower discharge rate after the wait cue during reward trials. B and C: average discharge rates of neurons with higher (24 neurons, 35 plots) and lower (23 neurons, 28 plots) activity in reward trials than in sound trials, but with no difference between aversion and sound trials. D: activity of a caudate neuron in monkey DA that was modulated selectively by aversion outcome just after the appearance of the outcome instruction. E and F: average discharge rates of neurons with higher (5 neurons) and lower (one neuron) activity in aversion trials than in sound trials, but with no difference between reward and sound trials. G: activity of a caudate neuron in monkey AI that was modulated selectively by sound outcome. H and I: average discharge rates of neurons with higher (5 neurons) and lower (one neuron) activity in sound trials than in other outcome trials. Circles and triangles represent data from monkeys DA and AI, respectively. Filled symbols in C, E, and H are the example neurons in A, D, and G, respectively.
reward preference was consistent across task epochs in most neurons (39/43). Reward outcome–dependent modulation was frequently observed during every task epoch after the outcome instruction (Fig. 8, A and B). In contrast, only a small number of neurons were selectively modulated during trials with aversion (6/70, Fig. 7, D–F) or sound outcome (6/70, Fig. 7, G–I). Discharge rates of 3 neurons were modulated by both reward and aversion outcomes at single task epochs. Thus after outcome instruction, striatal neurons maintained the instructed outcomes and modulated their behavioral response-related activity in an outcome-selective manner, especially the reward outcome.

Does the neuronal activity reflect behavioral measures of task performance?

To test whether neuronal activity reflects behavioral measures of task performance, we examined whether the variation of task start time was correlated with neuronal discharge rates before outcome instruction by using multiple regression analysis. In a neuron that exhibited a burst of discharges at the lever depression (Fig. 3D), average discharge rates in each trial were plotted against task start time in Fig. 9A. The regression analysis showed a significant regression slope of discharge rates with TLR (slope = 5.18, P < 0.001, R² = 0.316) but not with task start time (slope = -0.01, P = 0.61). The analysis of 44 neurons that were activated before outcome instruction revealed significant modulation of activity in 21 neurons by reward history (TLR) but not by task start time, in 6 neurons by both (Fig. 9B). The task start time influenced only a small number of neurons modulated by TLR (Fig. 9C). Thus the discharge rates of a subset of striatal neurons were modulated by reward history, but not by the behavioral performance, before the behavioral outcome was instructed.

We also examined the influences of behavioral measures on neuronal discharge rates after the outcome instruction, such as reaction time for the lever-release responses. The analysis of 70 neurons active after the outcome instruction revealed that discharge rates of only 9 neurons had significant regression slopes with the lever-release reaction time (Supplemental Table 1).

FIG. 8. Modulation of neuronal activity by the instructed outcome of current trials. Percentages of neurons modulated by reward, aversion, and sound outcomes during task epochs in monkey DA (A) and monkey AI (B). C: pie chart of the number of neurons that were modulated by current outcomes during the periods after the outcome instruction. Neurons modulated by both reward and aversion outcomes in single task epochs are indicated as “Reward and Aversion.” Neurons modulated by reward and aversion outcomes at different task epochs are represented with dark gray. NO indicates neurons without modulation by current outcomes.

Location of recorded neurons in the caudate nucleus and putamen

The recording sites of 163 neurons in the caudate nucleus and putamen of the two monkeys were histologically reconstructed (Fig. 10). During epochs before outcome instruction, 18 and 26 neurons showed increased discharge rates in the caudate nucleus and putamen, respectively (Fig. 10A). Neurons with increased discharge rates just before occurrence of the outcome instruction (triangles) were more frequently observed in the caudate nucleus (15/18) than in the putamen (3/26, Fisher’s exact probability test; P < 0.001). In contrast, neurons that showed an increase in discharge rate during the lever depression to initiate trials (squares) were more frequently observed in the putamen (19/26) than in the caudate nucleus (1/18, P < 0.001). Neurons that exhibited discharges during the wait cue period for the outcome instruction (Fig. 4A) were observed both in caudate nucleus (9/18) and putamen (7/26). About half of the neurons active before the outcome instructions (24/44) had discharges with significant positive or negative regression slopes with TLR (Fig. 4C). Thus the history-based processing occurred in both caudate nucleus and putamen.

After the outcome instruction (Fig. 10B), neurons with increased discharge rates just after occurrence of the outcome instruction were more frequently observed in the caudate nucleus (18/23) than in the putamen (6/47, P < 0.001). In contrast, neurons showing an increase in discharge rate during the lever release were more frequently observed in the putamen (42/47) than in the caudate nucleus (15/23, P = 0.022).
Modulation of neuronal activity by the outcomes of current trials after the outcome instruction occurred not only in the caudate nucleus (23/23) but also in the putamen (35/47). Thus caudate nucleus neurons had a tendency to be predominantly activated just before and after the outcome instruction, whereas putamen neurons were more active during lever depression and release movements.

**DISCUSSION**

We found that a subset of projection neurons in the caudate nucleus and putamen carry signals related to coding of reward outcome of forthcoming behavioral responses based on the reward history in the last several trials. The discharge rates of these neurons increased or decreased monotonically across trials in parallel with reward probability of the current trials. Reward history–dependent discharges of most neurons did not show significant regression slopes with behavioral measures of task performance, such as task start times. Once the behavioral outcome was instructed, the number of neurons displaying history-based coding of reward declined, and those encoding current outcome, especially for reward, increased. On the other hand, the history of aversion outcome influenced the activity of only a small number of neurons, and it had a weak impact on task performance. Although we used a beep sound as one of behavioral outcomes that selectively activated very limited number of neurons, the valence of the sound is not absolutely obvious: neutral event as we designed, an aversive event, or a cue for another salient event.

**History-based coding of reward outcome**

Rewards experienced as an outcome of previously attempted actions have strong influences on choosing subsequent actions (Thorndike 1898). In this study, the monkeys changed their behavioral responses depending on the probability of reward at current trials based on the recent history of reward (Fig. 2 and Table 3) after extensive periods of training. If the outcomes of taking the same actions vary among appetitive, aversive, and other stimuli, as occurred in the present study, estimation of when the next reward trials come becomes critical (i.e., the reward value of the current state). We found that the activity of a subset of neurons in the caudate nucleus and putamen could encode incrementing or decrementing reward probability across trials (Figs. 3B and 5). The reward history–based modulation of striatal neuron activity seemed to reflect estimated probability of reward (Fig. 3B). Alternatively, it might simply represent the number of trials since the last reward trial (TLR). This was not likely, however, because neurons in this study exhibited either gradually increasing or decreasing activity depending on TLR, unlike neurons representing numerical numbers in the parietal cortex (Nieder 2005; Sawamura et al. 2002).

In some neurons, discharge rates at the first trials after reward trials (TLR = 1) were much higher than those in other later trials (Fig. 5B). It is possible that this type of activity may also be related to the decrease in reward values specifically at the first trials following reward trials or preceding reward (Simmons and Richmond 2007). On the other hand, it might also be related to other factors such as the schedule process across trials, marking the start and end of a long-term schedule for reward. Although neuronal activity related to reward prediction across trials was reported in other brain areas (Hikosaka and Watanabe 2004; Ichihara-Takeda and Funahashi 2006; Shidara and Richmond 2002), the activity found in the present study was different from that in other brain areas, in that the presence and absence of reward in current trials was explicitly instructed before actions were initiated in these previous works, whereas in this study probability of individual out-

**FIG. 10.** Histological reconstruction of recording sites of neurons in the caudate nucleus and putamen in 2 monkeys. Recording sites of all 163 recorded neurons are plotted. A: neurons modulated by histories of reward (TLR; red), aversion (TLA; blue), and sound (TLS; yellow) are represented that show significant discharges to lever depression (○), or just before outcome instruction (△), or wait cue (□) before outcome instruction. Neurons never modulated by the histories were represented by black. B: neurons modulated by reward (red), aversion (blue), and sound (yellow) outcomes are represented that show significant discharges to lever release (○), or just after outcome instruction (△), or wait cue (□) after outcome instruction. Neurons never modulated by the current outcomes were represented by black. Significant discharges occurring during multiple task epochs were overwritten.
comes could be estimated only implicitly based on their histories in the last several trials.

Striatal neurons might encode reward values of task cues and intended actions in terms of magnitude, kinds, and probability of rewards (Cromwell and Schultz 2003; Hassani et al. 2001; Kawagoe et al. 1998; Samejima et al. 2005). Samejima et al. (2005) showed that the discharge rates of striatal neurons represent reward value of free-choice action (action value; probability × volume for each option) estimated by previously performed actions and their outcomes in the last several trials. This supported the reinforcement learning models of the basal ganglia (Doya 2000; Houk et al. 1995; O’Doherty et al. 2004), which proposed that action values represented in the striatum are used for the selection of the highest-value action and that the action values are updated by dopamine signals conveying reward prediction errors. Present observation of history-based coding of forthcoming behavioral outcomes might correspond to representation of state values in the reinforcement learning theories (Sutton and Barto 1998), which also play a major role in reward-oriented adaptive action selection. Dopaminergic neurons have been shown to carry signals of reward prediction errors (Bayer and Glimcher 2005; Fiorillo et al. 2003; Morris et al. 2004; Nakahara et al. 2004; Satoh et al. 2003). In addition, dopaminergic neurons signal errors of reward prediction estimated by the number of preceding unrewarded trials following a reward trial (Nakahara et al. 2004).

Coding of current outcomes of behavioral responses

After an outcome of behavioral responses was instructed, neuronal discharge rates were modulated by the types of outcome, especially reward, in a large number of presumed projection neurons (Figs. 7 and 8), consistent with previous reports (Hollerman et al. 1998; Kawagoe et al. 1998). Significantly fewer neurons (n = 9) were influenced by aversion in their responses to the task events of outcome instruction, wait cue, and lever-release responses (Fig. 8), except in monkey DA, where a considerable number of neurons were modulated by the instruction of aversion outcome (5/16, 31%, Supplemental Table 1). The percentage of the presumed projection neurons modulated by instruction of aversion outcome seemed to be similar to that of 317 tonically active neurons (TANs) in the striatum observed previously in the same monkey (25%; Yamada et al. 2004). TANs responded selectively either to outcome instruction or to lever release, whereas in projection neurons, discharge rates increased in relation to every task event not only before but also after outcome instructions. Projection neurons showed reward preference throughout task trials. Thus there was a clear difference of outcome-dependent activity between projection neurons and TANs as a population. Projection neurons in monkey AI might have a tendency to be influenced less by current aversion outcome than those in monkey DA (2/34 in monkey AI and 7/39 monkey DA, P = 0.16). This might be because monkey AI made behavioral responses very quickly at trials with every outcome instruction (Table 2). Thus the striatal projection neurons seemed to play a role in detection and discrimination of behavioral outcomes and to signal behavioral responses in a reward-dominant manner. Dominant modulation of neuronal activity by reward outcome over aversion outcomes of behavioral responses was also observed in the lateral prefrontal cortex of monkeys (Kobayashi et al. 2006), whereas the orbitofrontal cortex encodes relative preference of reward and aversion outcomes (Hosokawa et al. 2007).

One concern regarding the different influences of reinforcers on striatal neuron activity might be that, even if “aversion outcome” was instructed, airpuffs were actually delivered in only a small percentage of trials (a few times in a day) when lever-release responses were delayed. In other words, monkeys could avoid receiving airpuffs by quicker responses, and thus the instruction of “aversion outcome” might have had a smaller impact than that of “reward outcome” on task performances, although TANs previously observed in the same monkeys responded to instruction of aversion outcome as well as reward outcome (Yamada et al. 2004). Further study is necessary to clarify how much striatal projection neurons, particularly in its dorsal part, are sensitive to aversion outcome, while the functional role of ventral striatum in encoding aversion outcome was revealed (Setlow et al. 2003).

Influence of aversion history on activity of striatal neurons

Reward history–dependent coding was dominant over aversion history–dependent coding in the striatum (Fig. 4). There are several potential origins for this dominance of reward history. A simple one is the weaker impact of aversion history on task performance compared with reward history (Table 3), based on the actual, rather small number of airpuffs received, as described earlier for outcomes. However, the behavioral measures of task performance, such as task start time, influenced neuronal discharge rates in only a small number of striatal neurons in our sample (Fig. 9). Another possibility is that it was more important for monkeys to predict when the next reward would be available to obtain it, than to predict the next aversion trials to avoid it (Fig. 2, Table 3). Reward history–dependent coding may be dominant because aversion trials, which could be regarded as costs to obtain reward, had a relatively weak impact on neuronal discharge rates recorded in this study. If aversion trials had a stronger impact with airpuffs occurring every time at aversion trials (Blazquez et al. 2002), the percentage of neurons with aversion history–based coding might have been higher. Decision making depends on the balance between cost and reward and on relative preferences among different values of reward. Medial frontal cortex (Walton et al. 2002, 2003) and orbitofrontal cortex (Padoa-Schioppa and Assad 2006; Tremblay and Schultz 1999) seem to be involved in these processes. The striatum may also play major roles in valuation, decision, and selection of actions (Cromwell and Schultz 2003; Doya 2000; Graybiel et al. 1994; Houk et al. 1995; Kawagoe et al. 1998; O’Doherty et al. 2004; Samejima et al. 2005) in concert with processing in the frontal cortical areas through the cortico-basal ganglia loop circuits.

Functional implication of history- and instruction-based coding of forthcoming behavioral outcomes

The present study revealed that presumed projection neurons in the striatum encode behavioral outcomes, especially reward, in two different manners. Activity of a subset of neurons in the caudate nucleus and putamen encoded incrementing or decrementing reward probability across trials. The reward history–based signals may serve to guide behavior to reward by brisk
responses and low rate of errors when reward is impending. When reward is not expected, however, the behavioral responses are sluggish and erroneous (Shidara et al. 1998). This could underlie response bias, the behavioral observation that subjects are biased toward selecting one particular response over another (Lauwereyns et al. 2002; Watanabe et al. 2001). Once an outcome of the current trial became evident, striatal neurons selectively represented the current outcomes before, during, and after behavioral responses. The coding of current behavioral outcomes might play a role both in detecting and in maintaining the instructed outcomes. On the other hand, activity during behavioral responses (lever release) was often late in relation to response onset and seemed too late to play a major role in initiating the responses, such as in Fig. 6A. What could be the roles played by the coding of current response outcomes during and after behavioral responses? A conceivable role is to monitor or provide feedback on recently executed actions and their outcomes by providing downstream brain areas with the signals to evaluate the executed actions and evolve them for the next trials.

Previously, we showed that presumed cholinergic interneurons (TANs) in the caudate nucleus respond to stimuli associated with motivational outcomes, whereas TANs in the putamen are more related to movement-eliciting signals (Yamada et al. 2004). Consistent with this evidence, presumed projection neurons in the caudate nucleus had a tendency to be predominantly activated surrounding the outcome instruction, whereas those in the putamen were more active during lever depression and release movements. This supports the view that spatially segregated processing of cognition- and action-related signals by neuronal circuits occurs within the dorsal striatum. On the other hand, the local striatal circuits such as those in the putamen processing lever depression and release signals could also become specialized dynamically to process information about past trials before the outcome instruction and about current outcomes after their explicit instruction.

Both the history- and current instruction–based codes of behavioral outcomes might be neural substrates for action planning and learning in the dorsal striatum. Impairments in the coding may lead to deficits in predictive action planning for behavioral outcomes might be neural substrates for action planning for past trials before the outcome instruction and about current outcomes after their explicit instruction.

Both the history- and current instruction–based codes of behavioral outcomes might be neural substrates for action planning and learning in the dorsal striatum. Impairments in the coding may lead to deficits in predictive action planning for a distant goal (Dickinson and Balleine 1994) or habit formation (Yin and Knowlton 2006), which might underlie some of the core symptoms of Parkinson’s disease (Marsden 1984; Soliveri et al. 1997).

ACKNOWLEDGMENTS

We thank T. Minamimoto and K. Samejima for valuable comments; K. Yagi, A. Nishino, and M. Kawata for advice concerning statistical analyses; and R. Sakane for technical assistance.

GRANTS

This research was supported by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology, Japan to M. Kimura and by a Grant-in-Aid for Japan Society for the Promotion of Science Fellows to H. Yamada.

REFERENCES


