RESEARCH ARTICLE

Inactivation of the putamen selectively impairs reward history-based action selection

**Manabu Muranishi · Hitoshi Inokawa ·
Hiroshi Yamada · Yasumasa Ueda · Naoyuki Matsumoto ·
Masanori Nakagawa · Minoru Kimura**

Abstract Behavioral decisions and actions are directed to achieve specific goals and to obtain rewards and escape punishments. Previous studies involving the recording of neuronal activity suggest the involvement of the cerebral cortex, basal ganglia, and midbrain dopamine system in these processes. The value signal of the action options is represented in the striatum, updated by reward prediction errors, and used for selecting higher-value actions. However, it remains unclear whether dysfunction of the striatum leads to impairment of value-based action selection. The present study examined the effect of inactivation of the putamen via local injection of the $GABA_A$ receptor agonist muscimol in monkeys engaged in a manual reward-based multi-step choice task. The monkeys first searched a reward target from three alternatives, based on the previous one or two choices and their outcomes, and obtained a large reward; they then earned an additional reward by choosing the last rewarded target. Inactivation of the putamen impaired the ability of monkeys to make optimal choices during third trial in which they were required to choose a target different from those selected in the two previous trials by updating the values of the three options. The monkeys normally changed options if the last choice resulted in small reward (lose-shift) and stayed with the last choice if it resulted in large reward (win-stay). Task start time and movement time during individual trials became longer after putamen inactivation. But monkeys could control the motivation level depending on the reward value of individual trial types before and after putamen inactivation. These results support a view that the putamen is involved selectively and critically in neuronal circuits for reward history-based action selection.

Manabu Muranishi and Hitoshi Inokawa contributed equally to this work.

M. Muranishi (✉) · H. Inokawa · H. Yamada · Y. Ueda ·
N. Matsumoto · M. Kimura (✉)
Department of Neurophysiology, Kyoto Prefectural University
of Medicine, Kawaramachi-Hirokoji, Kamigyo-ku,
Kyoto 602-8566, Japan
e-mail: muranishi.manabu@jp.panasonic.com

M. Kimura
e-mail: mkimura@lab.tamagawa.ac.jp

M. Nakagawa
Department of Neurology, Kyoto Prefectural University
of Medicine, Kawaramachi-Hirokoji, Kamigyo-ku,
Kyoto 602-8566, Japan

H. Yamada
Center for Neural Science, New York University,
New York, NY, USA

M. Kimura
Brain Science Institute, 6-1-1,
Tamagawa Gakuen, Machida, Tokyo 194-8610, Japan

Introduction

Fundamental to decision-making is the ability to use past experience to select the best course of action among competing alternatives. In reinforcement learning theories, the problem of finding an optimal action in an uncertain environment is solved based on value functions representing the expected sum of future rewards for particular states or actions (Sutton and Barto 1998). The striatum is known as a key site involved in multiple cortico-basal

ganglia loop circuits including the motor loop through the putamen, oculomotor loop through the caudate nucleus, anterior cingulate loop through the ventral striatum (Alexander et al. 1986; Middleton and Strick 2000). The basal ganglia systems have been suggested to play a major role in action (DeLong et al. 1986; Desmurget and Turner 2008; Nambu 2008), purposeful behavior (Hikosaka et al. 2000; Kimura et al. 2004), and habit learning (Graybiel 2008; Tricomi et al. 2009; Ashby et al. 2010) through the integration of specific cortical inputs and dopaminergic modulatory inputs. In addition, a growing body of evidence suggests that the striatum adaptively encodes values of action options (action value) (Samejima et al. 2005; Hikosaka et al. 2006; Lau and Glimcher 2008) and of chosen actions (chosen value) (Pasquereau et al. 2007; Lau and Glimcher 2008). The encoded values are updated by reward prediction error signals provided by midbrain dopaminergic neurons (Schultz et al. 1997; Hollerman et al. 1998; Fiorillo et al. 2003; Satoh et al. 2003; Morris et al. 2004). In the reinforcement learning model of the basal ganglia, the value signals are mediated by the striatum (Houk et al. 1995; O'Doherty et al. 2004), whereas the cortico-basal ganglia loops mediate the comparison of values of action (Doya 2000). Specific involvement of the dorsal and ventral striatum in goal-directed and habitual responding (Balleine and O'Doherty 2010; Corbit and Janak 2010) and update of responding by outcomes (Ito and Doya 2009) have also been reported in rodents. However, it is still unknown how the value representation in the striatum contributes to action selection.

In the present study, we addressed this issue by blocking neuronal activity in the putamen via local injection of the $GABA_A$ receptor agonist muscimol into the putamen of monkeys engaged in a reinforcement-based multi-step choice task. The monkeys first searched for a target from three alternatives based on the histories of the last choices and their outcomes and obtained water as a reward (search epoch); they then could earn an additional reward by choosing the last rewarded target again on the basis of positive reinforcers (repetition epoch). After the putamen was inactivated locally by muscimol, the monkeys not only changed options if the last choice resulted in no reward (lose-shift) but also stayed with the last choice if it was rewarded (win-stay) normally. However, the rate of non-optimal choices increased at the third trial following two successive no reward choices where the monkeys chose an option already tried at the first choice. To make an optimal choice at the third trials, it was necessary for the monkeys to update values of individual options based on the previously tried two options and their outcomes and to choose highest-value options. Therefore, the specific effects of inactivation suggested pivotal roles of the putamen in reward history-based value update and action selection. Although the

motivation to work for reward may have declined because the time from the start cue to the initiation of trials increased, monkeys could control the motivation level depending on the reward value of individual trial types similar to the choices before putamen inactivation.

Materials and methods

Animals and surgery

Two female Japanese monkeys (*Macaca fuscata*; monkey TN, 5.8 kg and monkey YO, 6.0 kg) were used. All surgical and experimental procedures were approved by the Animal Care and Use Committee of Kyoto Prefectural University of Medicine and were in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals. Four head-restraining bolts and one stainless-steel recording chamber were implanted on the monkey's skull using standard surgical procedures. The monkeys were sedated with ketamine hydrochloride (10 mg/kg, i.m.) and then anesthetized with sodium pentobarbital (Nembutal; 27.5 mg/kg, i.p.). Supplemental Nembutal (10 mg/kg, 2 h, i.m.) was given as needed. A rectangular chamber (25 × 37 × 20 mm) was positioned on the left cerebral cortex at an angle of 45° under stereotaxic guidance to monitor the activity of putamen neurons and to insert the needle for injection of muscimol, as described below.

Behavioral task

To study how the putamen is involved in value- and task-strategy-based action selection, we employed a behavioral task for monkeys to make multi-step choices of one target from three alternatives for rewards. The monkeys were trained to sit in a primate chair facing a small panel placed 21 cm in front of their faces. Five LEDs were embedded on the panel: a small rectangular start button with a green light-emitting diode (LED) (start LED, 14 × 14 mm) at the bottom, 3 target buttons with green LEDs (target LEDs, 14 × 14 mm) in the middle row, and a small red LED (GO LED, 4 mm diameter) just above the center push buttons (Fig. 1a). Individual trials were initiated by illumination of the start LED. The monkeys depressed the illuminated start button with their right hand. When the monkeys continued to hold the button for 800 ms, the start LED was turned off and three target LEDs and a GO LED turned on simultaneously. The GO LED turned off, if the monkeys kept depressing the start button for another 50 ms. They then released the start button and depressed one of 3 illuminated target buttons (N1 trials). One of the 3 targets was associated with large reward, while the other 2 were
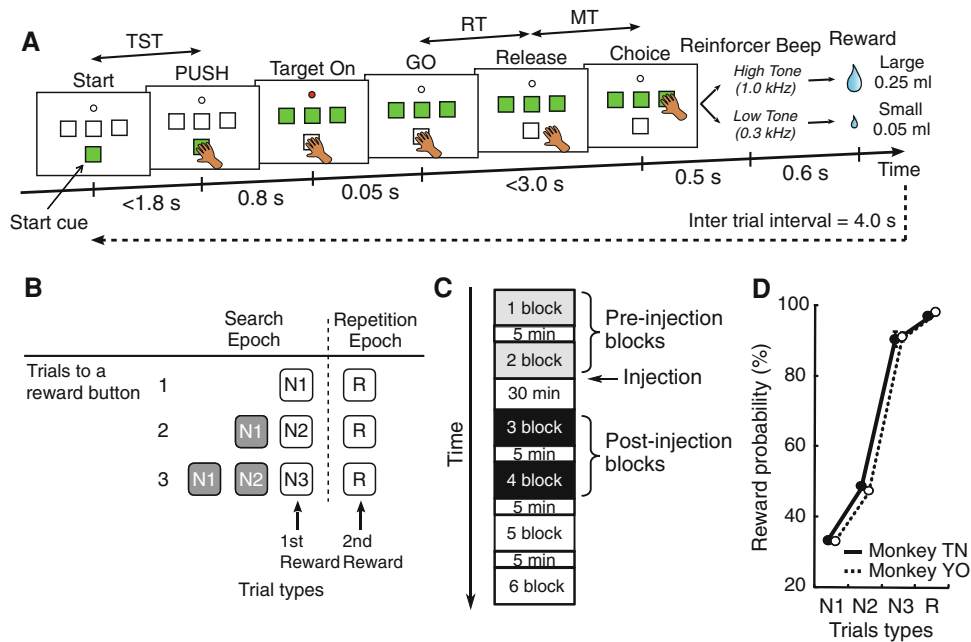
Fig. 1 Behavioral task and performance. a Illustration of sensorimotor events that occurred during single trials. TST, RTGO, and MT are task start time, reaction time to GO, and the time from release of the start button to depression of the target button, respectively. b Trial types during the search epoch and repetition epoch. *Gray* and *white rectangles* represent non-rewarded and rewarded trials, respectively.

c One block consisted of 12 trials for each trial type, and 6 blocks were performed per day. The first and second blocks were pre-injection blocks, and the third and fourth blocks were post-injection blocks. d Average reward probabilities during 4 types of trials in all pre-injection blocks for monkey TN (*black line*) and monkey YO (*broken line*). *Error bars* represent SEM

small-reward targets. If a small-reward button was depressed, a beep sound with a low tone (300 Hz, 100 ms) occurred with a delay of 500 ms, and a small amount of reward water (0.05 ml) was delivered through a spout attached to the monkey's mouth. If the monkeys chose the small-reward button again in the second trial (N2), the third (N3) trial started after a low-tone beep and a small reward had been presented. If a large-reward button was depressed, a beep sound with a high tone (1 k Hz, 100 ms) occurred with a delay of 500 ms, and a large amount of water (0.25 ml) was delivered. We used separate LEDs for the target on (illuminations of 3 green targets and a small red "pre-GO" signal) and for GO signal (offset of the "pre-GO" LED). Reaction time, from GO signal onset to release of the hold button, measured the time for monkeys to initiate choices after decisions had been made based on the preceding target signal.

The high-tone and low-tone beep sounds served as positive and negative feedback, respectively. Once the monkeys found a large-reward button during the search trials, they could obtain additional rewards by choosing the same button in the following repetition trial (R). The start button and the three target buttons flashed at the same time for 100 ms to inform the animal of the end of a series of trials. The next series of choice trials began at 4.0 s after the flashing of target buttons with the large-reward button appearing at a random target location.

Thus, the trials in a single series of choices were divided into two epochs (Fig. 1b). The first epoch was the search epoch, in which the monkey searched for a large-reward button on a trial-and-error basis. While an optimal strategy was to choose the button not selected in the previous trials (*lose-shift* strategy), this strategy was not sufficient for N3 trials in which monkeys had chosen small-reward buttons during the last two successive trials; i.e., instead, they had to choose the one remaining button, but not the one selected in the N1 or N2 trials. Thus, it was required for monkeys to choose the highest-value option among three alternatives while updating values of individual options based on the history of choices and their outcomes. The second epoch was the repetition epoch in which the monkeys again chose the large-reward button found in the last trials during the search epoch (*win-stay* strategy). One block consisted of at least 12 trials for each trial type. Task performance was studied parametrically during six blocks (2 pre-injection blocks, 2 post-injection blocks, and 2 additional blocks) in a day (Fig. 1c). Although the monkeys consistently performed task after muscimol injection during two post-injection blocks, they sometimes stopped performing the task during subsequent blocks. Therefore, we used 2 blocks of trials for pre-injection data and two additional post-injection blocks of trials (the third and fourth blocks) after muscimol injection as.

Electrophysiological mapping and muscimol injections

After recovery from surgery, single-unit recordings were made to map the rostral, middle, and caudal parts of the putamen during performance of the behavioral task. We used epoxy-coated tungsten microelectrodes (Frederick Haer Company, Bowdoinham, ME) with an exposed tip of 15 μm and impedances of 2–5 MΩ (at 1 kHz). The neuronal activity was amplified and displayed on an oscilloscope using conventional electrophysiological techniques. Bandpass filters (50 Hz–3 kHz bandpass with a 6 dB per octave roll-off) were used. The action potentials of single neurons were isolated by using a spike sorter with a template-matching algorithm (MSD4; Alpha Omega; Nazare), and the duration of negative-going spikes was determined at a resolution of 40 μs. The onset times of the action potentials were recorded on a laboratory computer, together with the onset and offset times of the stimulus and behavioral events that occurred during the behavioral tasks. The electrodes were inserted through the implanted recording chambers and advanced by means of an oil-drive micromanipulator (MO-95; Narishige, Tokyo, Japan). To identify the topographical location of the putamen, we made recordings of multi- and single-unit activity through the course of the cerebral cortex dorsally, then the putamen and the globus pallidus ventrally in the middle and posterior levels of the putamen. These three structures show distinctive patterns of activity, such as very low background firing and infrequently occurring bursting discharges characteristic of striatal projection neurons, tonic-activity and tonically active characteristic of cholinergic interneurons, and very high frequency spikes characteristic of the globus pallidus (Yamada et al. 2004; Hori et al. 2009; Inokawa et al. 2010). For mapping the putamen, recordings were made from 35 locations of electrode penetrations in Monkey TN and from 15 locations in Monkey YO. Following injection of muscimol or saline in the putamen, neuronal spike activity was recorded by using a fine wire electrode (50 μm diameter) attached to the injection cannula to confirm that the injection sites were in the expected locations in the putamen.

The effects of muscimol and saline injections in the putamen on the task were studied after the completion of electrophysiological mapping of the putamen. Based on the effects of muscimol injection on task performance, the injection sites were separated into three parts: anterior level (3 mm anterior to the anterior edge of the anterior commissure, AC), middle level (3 mm posterior to the anterior edge of the AC), and posterior level (4–7 mm posterior to the anterior edge of the AC). The unilateral injections were made in the putamen (left hemisphere) contralateral to the arm used for button selection (right hand). Muscimol (5 μg/μl) or isotonic saline was injected locally in the putamen through

30-gauge cannula with a beveled tip which was connected by a fine polyethylene tube to a Hamilton syringe (5 μl). The injection speed was 0.25 μl/min, and the total injection volume was controlled by an electrically controlled injector (Baby Bee; Bioanalytical Systems, Inc., West Lafayette, USA). The injection volume was 2.0 or 3.0 μl at each site. The muscimol injection was expected to inactivate striatal neurons located around 2 mm in diameter based on the simultaneous recording of neuronal activity and muscimol injection (Shima and Tanji 1998). Post-injection blocks started 30 min after the injections were completed, because the effects of muscimol on task performance appeared at about 30 min (Shima and Tanji 1998; Sawaguchi and Iba 2001).

Data analysis

Three types of behavioral parameters were defined: task start time (TST) from illumination of start cue to depression of the start button, reaction time (RT) from onset of GO signal to the release of the start button, and movement time (MT) from the release of the start button to depression of the target button. These parameters served as motor indices. They were quantitatively compared before and after muscimol injection by using ANOVA ($P < 0.05$). To evaluate reward probability-dependent changes of motivation, the correlation coefficients between reward probabilities and TSTs were deemed to be statistically significant at $P < 0.05$. Speed of arm movement was evaluated by assessing movement times for each target button.

Choice data were pooled and compared between pre-injection (first and second) blocks and post-injection (third and fourth) blocks by the use of Fisher's exact probability test with the threshold for statistical significance set at $P < 0.05$. The effects of muscimol injection on task strategy were evaluated by examining the choices with valid (lose-shift and win-stay) and invalid (lose-stay and win-shift) strategies before and after the injection on N2, N3, and R trials (Fisher's exact probability test, $P < 0.05$). Optimal choices for value-based decision-making were defined as choosing higher-value options among three alternatives. In the N3 trial, the monkeys made one of three types of choices: choice of the button tried at N1 trials (non-optimal), choice of the button tried at N2 trials (non-optimal), and choice of the one remaining button (optimal).

Histological examination

After all behavioral experiments were completed, small electrolytic lesions were made at 20 locations along 10 selected electrode tracks in the putamen while monkeys were quietly sitting on the primate chair. In many cases, micro-lesions were made at the border between the

putamen and the external segment of the globus pallidus using the neuronal discharge properties as a guide. Direct anodal current (20 μA) was passed for 30 s through tungsten microelectrodes. The monkeys were deeply anesthetized with an overdose of pentobarbital sodium (90 mg/kg, i.m.) and perfused with 4% paraformaldehyde in 0.1 M phosphate buffer through the left ventricle. Coronal sections of the striatum, 50 μm in thickness, were stained with cresyl violet. The tracks of the microelectrode and injection needle through the putamen were reconstructed on the histology sections using the electrolytic lesion marks as reference points, and the locations of the muscimol and saline injections were identified.

Results

A total of 17 muscimol injections (10 in monkey TN, 7 in monkey YO) and 9 saline injections (3 in monkey TN, 6 in monkey YO) were made into the putamen of two

hemispheres of two monkeys (Table 1). The locations of all injections are summarized in Fig. 2.

Effects of inactivation of the putamen on motivation to start trials for multi-step choices

We measured TST, as a conventionally used index for motivation to work for reward (Shidara et al. 1998; Watanabe et al. 2001; Lauwereyns et al. 2002; Satoh et al. 2003). Figure 3 plots the TSTs against reward probabilities of individual trial types. The TSTs were negatively correlated with the reward probabilities: shortest at highest probability (R trials) and longest at lowest probability (N1 trials). In both monkey TN and monkey YO, the TSTs after muscimol injection became significantly longer in all trial types than those before injection in the anterior, middle, and posterior levels of the putamen (ANOVA, injection effect, monkey TN: anterior level, $F_{1,3} = 29.85$, $P < 0.0001$; middle level, $F_{1,3} = 158.0$, $P < 0.0001$; posterior level, $F_{1,3} = 96.24$, $P < 0.0001$, monkey YO: anterior level,

Table 1 Summary of the injection sites and effects

Monkey/site		Drug	Volume	Injection number	Distance from AC	Non-optimal choices at N3 trials	MTs at N1 trials
Monkey TN							
Anterior level	a	Muscimol	3 μl	1	+3.0 mm	NS	NS
	b	Muscimol	2 μl	1	+3.0 mm	NS	NS
	c	Saline	2 μl	1	+2.0 mm	NS	NS
Middle level	d	Muscimol	2 μl	1	±0.0 mm	↑	NS
	e	Muscimol	2 μl	1	±0.0 mm	NS	↑
	f	Muscimol	3 μl	1	−1.0 mm	↑	NS
	g	Muscimol	3 μl/2 μl	1/1	−2.0 mm	NS/↑	↑/↑
	h	Saline	2 μl	1	−2.0 mm	NS	NS
Posterior level	i	Muscimol/Saline	2 μl/2 μl	1/1	−5.0 mm	NS/NS	NS/NS
	j	Muscimol	2 μl	1	−6.0 mm	NS	↑
	k	Muscimol	2 μl	1	−6.0 mm	NS	↑
Monkey YO							
Anterior level	l	Saline	2 μl	1	+3.0 mm	NS	NS
	m	Muscimol	2 μl	1	+3.0 mm	NS	NS
	n	Saline	2 μl	1	+3.0 mm	NS	NS
	o	Muscimol	2 μl	1	+2.0 mm	NS	NS
Middle level	p	Muscimol	2 μl	1	−1.0 mm	↑	↑
	q	Muscimol	2 μl	1	−1.0 mm	NS	NS
	r	Saline	2 μl	1	−2.0 mm	NS	NS
	s	Muscimol	2 μl	1	−2.0 mm	↑	NS
	t	Saline	2 μl	1	−2.0 mm	NS	NS
Posterior level	u	Saline	2 μl	1	−6.0 mm	NS	NS
	v	Muscimol	2 μl	1	−6.0 mm	NS	↑
	w	Muscimol	2 μl	1	−6.0 mm	NS	↑
	x	Saline	2 μl	1	−7.0 mm	NS	NS

The direction of *arrows* indicates an increase of the value (Fisher's exact probability test). *NS*, statistically not significant; *AC*, anterior commissure
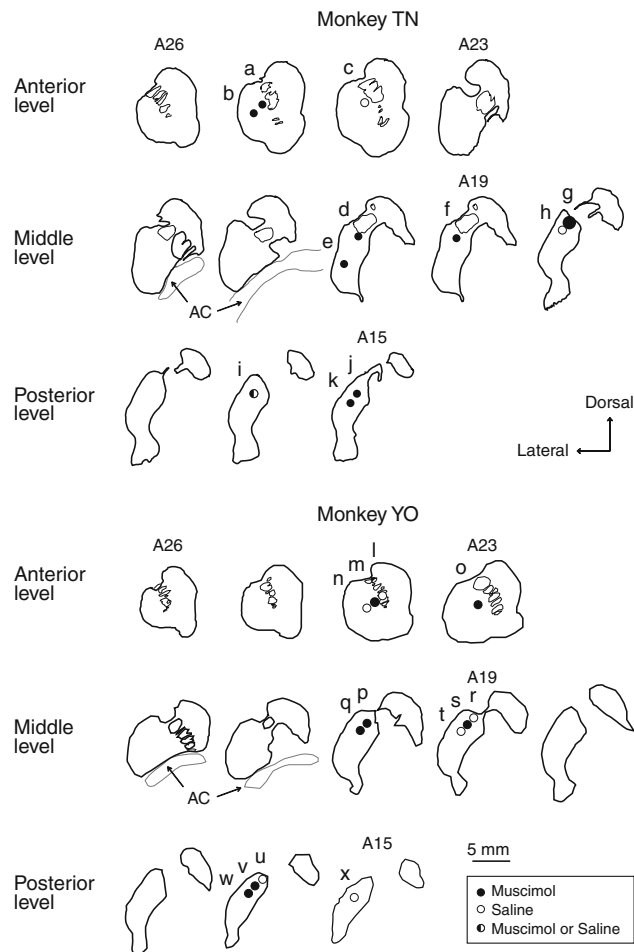
Fig. 2 Sites of injection of muscimol and saline. Sites of individual injections reconstructed from histology in monkey TN (*upper panel*) and YO (*bottom panel*). Symbol sizes indicate injection number (large circle denotes two injections, and small circles, single injection). Filled circles, open circles, and half-filled circle denote the injection sites of muscimol, saline, and muscimol or saline, respectively

$F_{1,3} = 16.52$, $P < 0.0001$; middle level, $F_{1,3} = 119.4$, $P < 0.0001$; posterior level, $F_{1,3} = 16.08$, $P < 0.0001$) (Fig. 3a). Notably, the negative correlation between TSTs and reward probabilities was maintained after muscimol injections for all injection sites. On the other hand, there was no consistent change in TSTs after saline injection (shortening after injection at the middle level of the putamen in monkey TN, lengthening after injection at the anterior and posterior level of monkey YO, and no significant change after the other injections) (Fig. 3b). These results indicate that inactivation of the putamen did not impair the processes of estimation of trial type-specific reward value and of reward value-dependent motivation to start individual choices for reward: i.e., there was a low level of motivation with low reward probability and a high level of motivation with high reward probability.

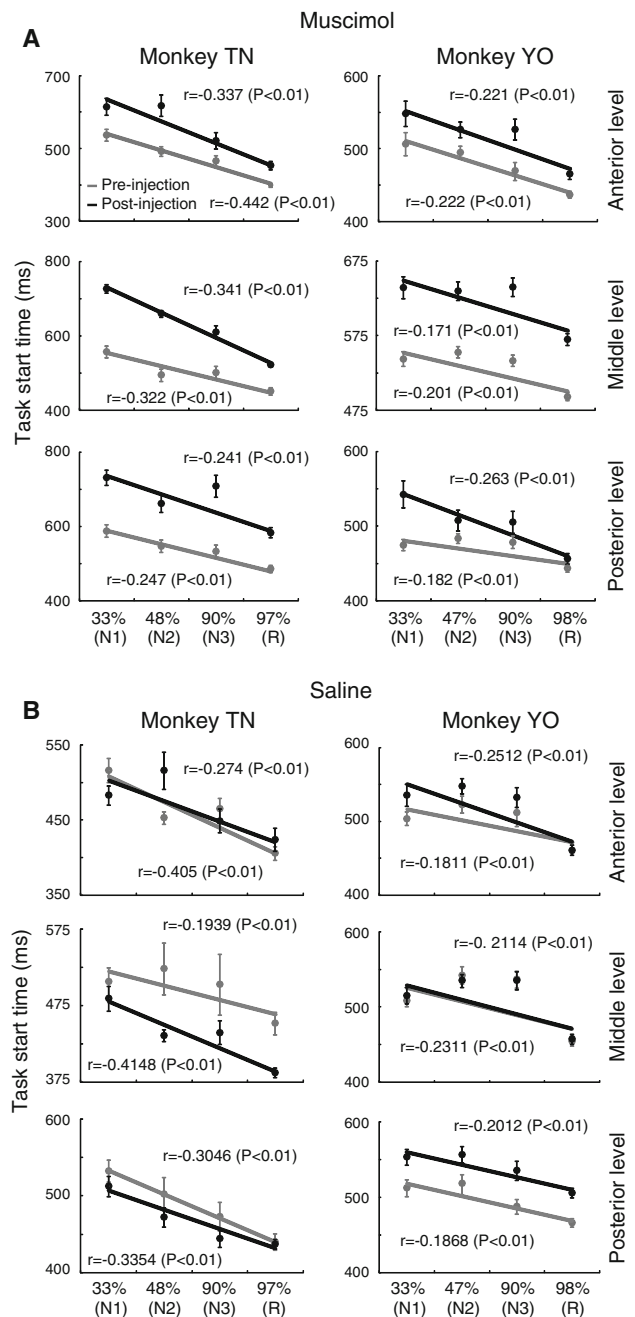It would be possible that reaction time would change depending on factors other than motivational level at the

Fig. 3 Negative correlation between task start time and reward probability for individual choices was maintained after muscimol injection. **a** Plots of task start time against reward probability before and after muscimol injections. The injections were made at the anterior, middle, and posterior level of the putamen in monkey TN (*left column*) and monkey YO (*right column*). **b** Same plots before and after saline injections. *Gray lines* denote task start times during pre-injection blocks, and *black lines* denote those during post-injection blocks. Data represent mean ± SEM. Regression lines are superimposed

start of trials, such as the number of response options and the number of previous choices necessary to remember and to decide on an optimal choice in current trials. However,

the TST, sum of reaction time and movement time became shorter as the number of previous choices to remember and to decide on an optimal choice increased (Fig. 3). In other words, TST was negatively correlated with number of response options. Thus, among possible factors influencing TST as a function of N1–N3, motivation to work for reward appeared to be the most critical one.

Inactivation of putamen impairs reward history-based action selection

Although task strategies (lose-shift and win-stay) were essential components for optimal performance of the task, they were insufficient in the case of N3 trials in which monkeys had chosen small-reward buttons during the last two successive trials. Monkeys had to choose the one remaining button, but not buttons chosen during the N1 or N2 trials. In other words, monkeys chose the highest-value option among three alternatives while updating values of individual options based on the histories of choices and their outcomes. Figure 4 shows the rate of choosing buttons not tried at immediately preceding choices (lose-shift rate) during the search epoch (Fig. 4a), and the rate of choosing the same button as in the last trials (win-stay rate) during the repeat epoch (Fig. 4b). In both of two monkeys, the very high lose-shift and win-stay rates were maintained after muscimol injection (Table 2). Thus, the monkeys could perform the multi-step choice task for rewards based on the lose-shift and win-stay strategy under local inactivation of the putamen.

As shown in the representative results in Fig. 5, the rate of non-optimal, small-reward choices increased selectively at N3 trials after muscimol injection (Fisher's exact probability test, $P < 0.05$). This was observed in both of the monkeys examined. The non-optimal choices occurred by choosing the button that had already been chosen (Fig. 5a, arrows) and resulted in small reward during the N1 trials. Thus, the choices were valid for the lose-shift strategy but were non-optimal for choosing the highest-value option. The rate of non-optimal choices in the N2 and R trials remained very low after muscimol injection (Fig. 5b). An increase in the non-optimal N3 choice rates occurred after muscimol injection in the middle anterior–posterior level of the putamen (Fig. 6a, $P < 0.05$, Fisher's exact probability test), whereas no significant change was evident after injections into the anterior and posterior levels (Fig. 2). When monkeys made a non-optimal N3 choice, they kept choosing until they got the large reward. After muscimol injection in the middle anterior–posterior level of the putamen, the large-reward target was reached within two additional trials in 93% of non-optimal N3 choices in Monkey TN (74% in one additional trial) and in 90% in Monkey YO (71% in one additional trial). Thus, the number of N3 trials increased after local inactivation of
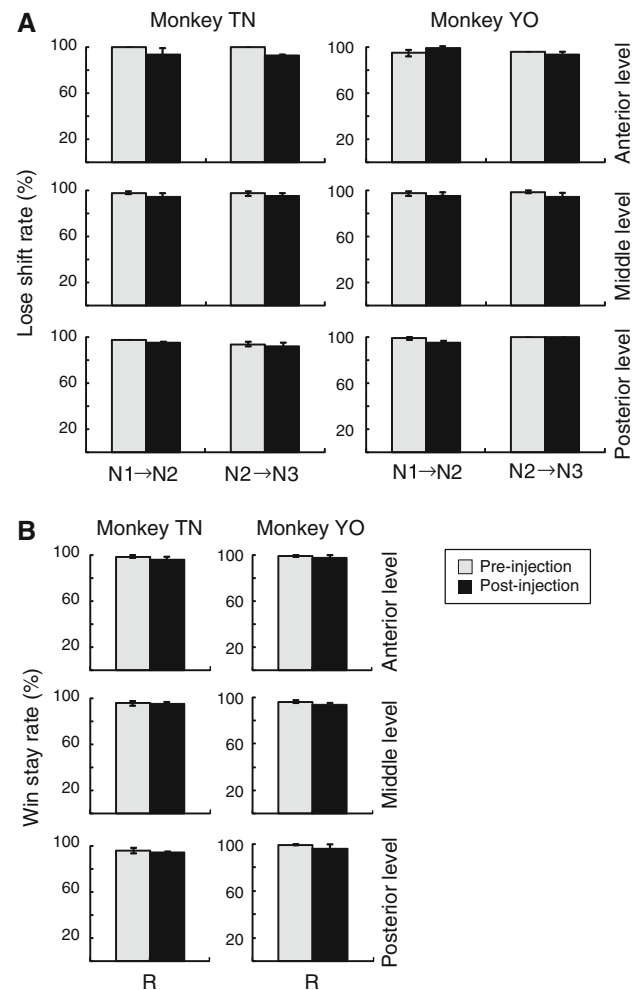

Fig. 4 Inactivation of the putamen did not impair lose-shift and win-stay strategies for multi-step choices. **a** Rate of lose-shift before and after muscimol injections in the anterior, middle, and posterior parts of the putamen in monkeys TN and YO. N1→N2 and N2→N3 indicate lose-shift rate from N1 to N2 trials and from N2 to N3 trials, respectively. **b** The rate of win-stay when the trials were switched from search to repetition trials (R). *Error bars* represent SEM

the putamen. Once this occurred, there were two or three N3 trials in a row, such as shown in Fig. 5a. Non-optimal choice rates in the N2 and R trials remained very low after each of the 17 muscimol injections (Fig. 6a). The rate of non-optimal choices in the N2, N3, and R trials did not change significantly following injections of physiological saline at any site in the putamen (Fig. 6b). Most of the non-optimal N3 choices occurred when the monkeys chose buttons that were already chosen in the N1 trials (Fig. 5a).

There would be two critically important components of lost functions after putamen inactivation related to the choice of the N1 buttons again during N3 trials. One is the working memory load: monkeys chose a different target from the last one selected (lose-shift) during search choices and the same target (win-stay) during repetition choices by

Table 2 *P* values for lose-shift and win-stay strategies

	Lose-shift			
	Monkey TN		Monkey YO	
	N1→N2	N2→N3	N1→N2	N2→N3
Anterior level	0.06	0.25	0.21	>0.99
Middle level	0.10	0.46	0.37	0.20
Posterior level	0.05	0.73	0.30	>0.99
	Win-stay			
	Monkey TN R		Monkey YO R	
Anterior level	>0.99		0.62	
Middle level	0.62		0.27	
Posterior level	0.17		0.08	

P values determined using Fisher's exact probability test
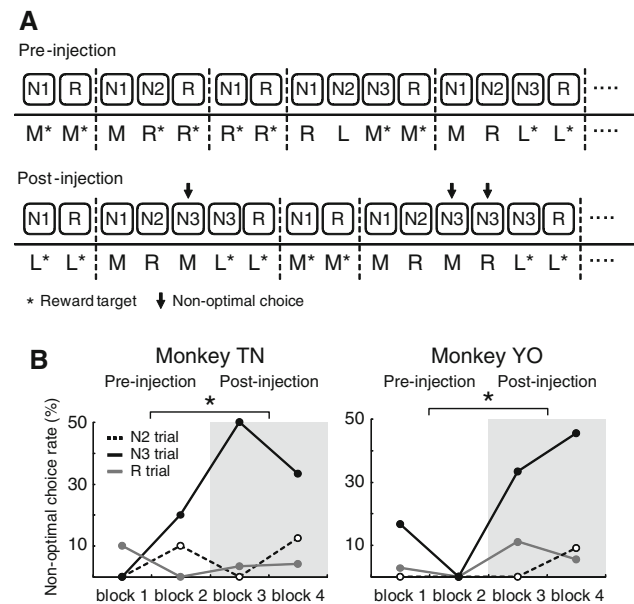

Fig. 5 Inactivation of the putamen at the middle level increases the non-optimal choice rate in N3 trials in the case of muscimol injection (specimen). **a** An example of action selection during the multi-step choice task before and after muscimol injection in monkey TN. Asterisks and arrows denote reward target and non-optimal choice, respectively. *Dashed lines* denote the end of one trial. L, M, and R denote for left, middle, and right target choice, respectively. **b** Example of changes in non-optimal choice rate in each trial type for monkey TN (*left column*) and monkey YO (*right column*). Comparisons were made between pre-injection blocks and post-injection blocks. *Broken lines*, *black lines*, and *gray lines* denote for inappropriate choice rate in N2 trials, N3 trials, and R trials, respectively. * *P* < 0.05 Fisher's exact probability test between pre- and post-injections

remembering the last choices (Fig. 4). But, in the N3 trials, they had to remember not only the last N2 choice but also the N1 choice. The other was an imperfect value-based choice. Monkeys can choose one reward target among
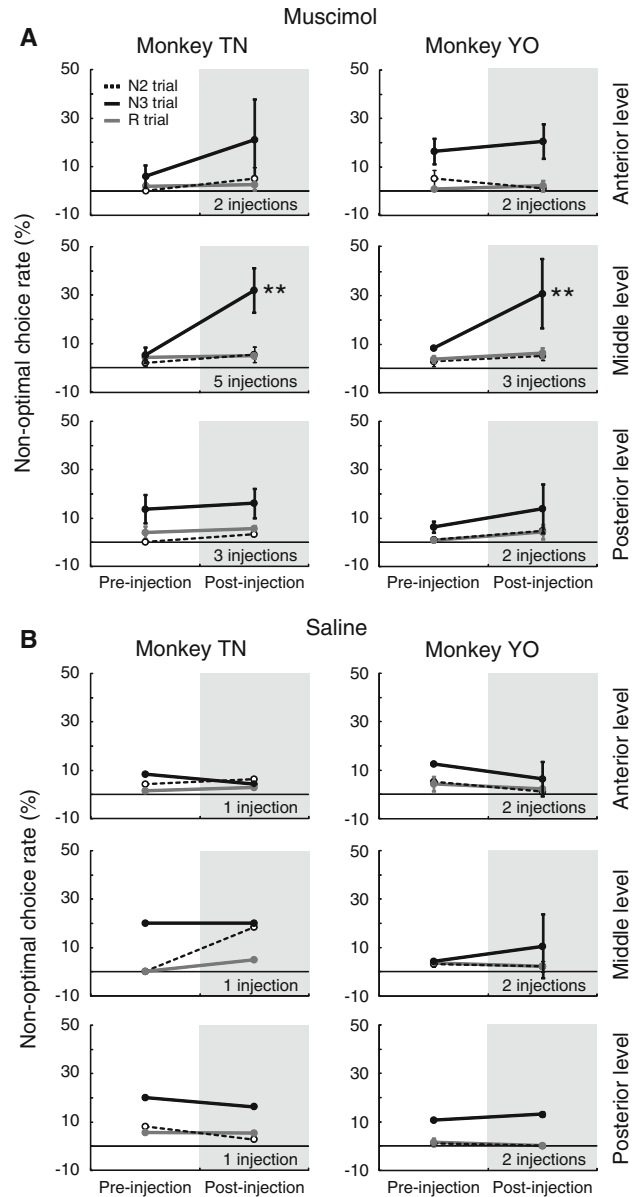

Fig. 6 Changes in the non-optimal choice rate for monkey TN (*left column*) and monkey YO (*right column*). **a** Muscimol injections. **b** Saline injections. Comparisons were made between pre-injection blocks and post-injection blocks. *Broken lines*, *black lines*, and *gray lines* denote the non-optimal choice rate in N2 trials, N3 trials, and R trials, respectively. Data represent mean ± SEM. ** *P* < 0.01 Fisher's exact probability test

three alternatives by updating the values of chosen targets depending on their outcomes: i.e., lowering after small reward and elevating after large reward. But, because working memory would provide knowledge of previously chosen options and their outcomes in the processes of the history-based value update and action selection, inactivation of local activity of the putamen in this study suggests composite functions of the putamen in decision-making and action selection.

Slowness of movement after inactivation of the putamen at middle and caudal putamen

We examined the effects of inactivation of the putamen on behavioral measures of task performance. Figure 7 shows movement times from release of the start button to depression of the target button during N1 trials. Movement times became longer after muscimol injection in the middle and posterior level of the putamen. The lengthening of movement times occurred for all 3 target buttons (left, middle, and right) (Fig. 7a). However, there was no significant change in movement times after injection in the anterior part of the putamen (Bonferroni correction, monkey TN: left target, $P = 0.09$; middle target, $P = 0.43$; right target, $P = 0.54$, monkey YO: left target, $P = 0.08$; middle target, $P = 0.25$; right target, $P = 0.33$). In control experiments with saline injection, there was no lengthening of movement times for any injection site (Fig. 7b). This observation was consistent with the previous results of inactivation of the striatum (Miyachi et al. 1997) and blockade of glutamatergic transmission in the globus pallidus (Kato and Kimura 1992).

Discussion

In the present study, we found three lines of evidence suggesting critical involvement of the putamen in reward history-based action selection. First, after the putamen was inactivated locally, the monkeys normally changed options if the last choice resulted in small reward (lose-shift) and stayed with the last choice if it was followed by large reward (win-stay). However, the rate of non-optimal choices increased at the third trials following two successive small-reward choices where the monkeys chose an option already tried at the first choice. At the third choices, monkeys had to update values of individual options based on the previously tried two options and their outcomes and to choose highest-value options. Therefore, the specific effects of inactivation suggested pivotal roles of the putamen in reward history-based value update and action selection. On the other hand, although non-optimal choices at N3 trials significantly increased after muscimol injection, the correct choice rate was still considerably higher (74% in monkey TN, 71% in monkey YO) than that of N2 trials (48% in monkey TN, 46% in monkey YO). This was probably due to the fact that inactivation by muscimol injection (2–3 µl, 5 µg/µl) covered limited areas of the putamen. Second, the effects of inactivation of the putamen on reward history-based action selection were especially strong at the middle rostro-caudal level, but were not significant at the rostral and caudal level. Third, reward value-dependent motivation to work for reward did not appear to be influenced by local inactivation of the putamen.
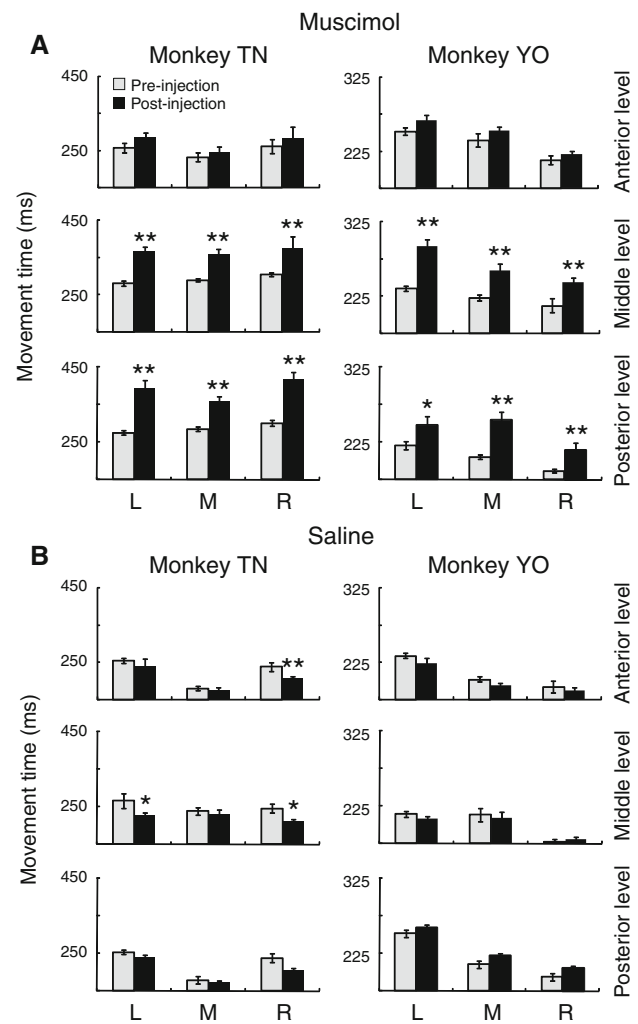

Fig. 7 Slow movement of task performance after inactivation of the putamen at the middle and posterior levels. **a** Movement time from release of start button to depression of left (L), middle (M), and right (R) button before and after muscimol injections during N1 trials for monkey TN (*left column*) and monkey YO (*right column*). **b** Data for saline injections (same paradigm as in previous panel). *Bars* graphs represent mean ± SEM. * and ** $P < 0.05$ and $P < 0.01$ in ANOVA

Brain circuit for reward history-based action selection and involvement of the striatum

Theories of reinforcement learning describe reward-based decision-making and adaptive choice of actions by estimating how the extent of the rewards a series of actions will yield (value function), and selecting the action by updating and comparing the value function of multiple alternatives based on the reward prediction errors (Sutton and Barto 1998). Midbrain dopaminergic neurons encode errors of reward expectation (Schultz et al. 1997; Satoh et al. 2003; Morris et al. 2004) as well as salience of events and motivation for actions (Redgrave et al. 1999; Satoh et al. 2003; Matsumoto and Hikosaka 2009). The frontal cortex (Matsumoto et al. 2003; Barraclough et al. 2004;

Daw et al. 2006), parietal cortex (Platt and Glimcher 1999; Sugrue et al. 2004), and basal ganglia (Lauwereyns et al. 2002; Samejima et al. 2005; Morris et al. 2006; Lau and Glimcher 2008) have been suggested to play a major part in value-based decision-making and choice behavior.

Neurons in the anterior cingulate cortex (ACC) of monkeys display modulation of activity related to the degree of reward expectancy estimated by previous experiences (Shidara and Richmond 2002) and to the rewards in previous trials (Seo and Lee 2007). Lesions of the ACC in monkeys do not impair reinforcement-guided choices immediately after errors but make the monkeys unable to sustain rewarded responses (Kennerley et al. 2006), suggesting critical involvement of the ACC in integrating information of chosen actions and their outcomes over time for guiding future actions. Lesions of the orbitofrontal cortex caused a deficit in stimulus selection but not action selection based on the previous reward experiences, in contrast with lesions of the ACC (Rudebeck et al. 2008). In our study, inactivation of the middle level of the putamen caused impairment of multi-step choices based on the action and reward history. This was in contrast to the fact that choices guided by a simple strategy of lose-shift and win-stay immediately following the choices remained intact (Fig. 4). Thus, these results support a view that the putamen, especially at the middle rostro-caudal level, plays a vital role in choices based on the action and reward history, which include integration and update of action and reinforcement information over time.

The motivation to work for reward may have declined after inactivation of the putamen, because the lengthening of TSTs was occurred after muscimol injection (Fig. 3). However, the monkeys could control the level of motivation depending on the reward value (probability) of individual choices: i.e., they were highly motivated (short TSTs) when the value of choices was high and vice versa (Fig. 3) after putamen inactivation. This suggested that motivational control of value-based choices is achieved mostly through other cortico-basal ganglia loop circuits, such as those involving the caudate nucleus and ventral striatum. It is unclear whether muscimol-induced lengthening of TSTs without significant change in reaction time to GO signal reflects a selective slowing of internally guided or triggered movements, because both TSTs and reaction time to GO signal are measures of triggered movements.

Working memory function

It could be argued that the deficits in reinforcement-guided choices after inactivation of the putamen are attributable to a general failure of working memory, which might compromise recall of the actions and outcomes experienced in previous trials. Although there is a mnemonic component in remembering the history of past actions and outcomes,

the results of previous studies of inactivation of neuronal activity and blockade of dopaminergic functions in the putamen cannot simply be ascribed to deficits in the process of remembering (Monchi et al. 2001; Coull et al. 2008; Kojima et al. 2009; Beck et al. 2010), in contrast with the results of studies in which the lateral prefrontal cortex was lesioned (Fuster 1991; Goldman-Rakic 1996). Vulnerability to working memory overload may be mediated by reduced activity of the prefrontal-limbic system (e.g., amygdala, hippocampus) (Monchi et al. 2001; Yun et al. 2010).

Matching behavior after negative and positive feedback (lose-shift and win-stay) was executed almost perfectly in this study without a significant influence of the putamen inactivation. However, inactivation of the putamen led monkeys to make errors in N3 trials as a result of choosing N1 buttons (Figs. 5, 6). Thus, the most critical functions that were lost after putamen inactivation were consistent with the reward history-based update of values of chosen options for action selection, part of which includes known components of working memory, such as short-term maintenance and manipulation of information (Baddeley and Hitch 1974).

Region-specific effects of inactivation on functions of the putamen

In the present study, inactivation at the middle rostro-caudal level of the putamen had a significant effect on choices based on the histories of previous choices and their outcomes. This part of the putamen receives dense projections from the medial frontal cortical areas, especially from part of the ACC that also innervates limbic basal ganglia circuits (McFarland and Haber 2000; Takada et al. 2001; Haber et al. 2006). Consistent with these cortico-striatal projections, accumulating evidence suggests critical involvement of the ACC in integrating information of chosen actions and their outcomes over time for guiding future actions (Kennerley et al. 2006; Rudebeck et al. 2008). The caudal region of the putamen receives projections predominantly from motor-related cortical areas (Flaherty and Graybiel 1995; McFarland and Haber 2000; Nambu et al. 2002). Inactivation of the middle and caudal part of the putamen induced slower movement in task performance (Fig. 7a), which is consistent with the predominant projections from motor and somatosensory cortical areas. Inactivation of the major target of the putamen, the globus pallidus, influences the kinematics of task movement (Kato and Kimura 1992; Desmurget and Turner 2008; Desmurget and Turner 2010).

Although a total of 17 locations of muscimol injection covered wide areas of the putamen in two monkeys, the effects of inactivation were still limited to the relatively

dorsal part of the putamen and the ventral part was not examined (Fig. 2). Thus, the present study did not necessarily test all possible functions of the putamen, but focused on reward-based evaluation and selection of actions. This was because recent studies on the striatum emphasize evaluative functions such as representation of values of actions and stimuli and outcomes (Kawagoe et al. 1998; Samejima et al. 2005; Lau and Glimcher 2008; Hori et al. 2009). Involvement of the limbic cortico-basal ganglia circuits through the ventral striatum is also suggested in reward-based action selection (Cardinal and Howes 2005; McCoy and Platt 2005; Nicola 2007; Ito and Doya 2009). Processing of values for decision and action selection in the putamen, caudate nucleus, and ventral striatum appear to depend on the value-specific inputs from wide cortical areas (Haber et al. 2006) and from midbrain dopaminergic neurons (Haber and Knutson 2010). Thus, the involvement of the putamen in reward history-based action selection which we found in this study seems to reflect a common aspect of the basic functions of the striatum and cortico-basal ganglia system, such as proposed by reinforcement learning models of the basal ganglia in value of actions are encoded in the striate projection neurons and updated by dopamine-mediated prediction error signals to select a series of actions expected to maximize rewards (Houk et al. 1995; Schultz et al. 1997; Sutton and Barto 1998; Doya 2000; O'Doherty et al. 2004).

References

Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu Rev Neurosci 9:357–381

Ashby FG, Turner BO, Horvitz JC (2010) Cortical and basal ganglia contributions to habit learning and automaticity. Trends Cogn Sci 14:208–215

Baddeley AD, Hitch GJL (1974) Working Memory. In: Bower GA (ed) The psychology of learning and motivation: advances in research and theory vol 8. Academic Press, New York, pp 47–89

Balleine BW, O'Doherty JP (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology 35:48–69

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

Beck SM, Locke HS, Savine AC, Jimura K, Braver TS (2010) Primary and secondary rewards differentially modulate neural activity dynamics during working memory. PLoS One 5:e9251

Cardinal RN, Howes NJ (2005) Effects of lesions of the nucleus accumbens core on choice between small certain rewards and large uncertain rewards in rats. BMC Neurosci 6:37

Corbit LH, Janak PH (2010) Posterior dorsomedial striatum is critical for both selective instrumental and Pavlovian reward learning. Eur J Neurosci 31:1312–1321

Coull JT, Nazarian B, Vidal F (2008) Timing, storage, and comparison of stimulus duration engage discrete anatomical components of a perceptual timing network. J Cogn Neurosci 20:2185–2197

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876–879

DeLong MR, Alexander GE, Mitchell SJ, Richardson RT (1986) The contribution of basal ganglia to limb control. Prog Brain Res 64:161–174

Desmurget M, Turner RS (2008) Testing basal ganglia motor functions through reversible inactivations in the posterior internal globus pallidus. J Neurophysiol 99:1057–1076

Desmurget M, Turner RS (2010) Motor sequences and the basal ganglia: kinematics, not habits. J Neurosci 30:7685–7690

Doya K (2000) Complementary roles of basal ganglia and cerebellum in learning and motor control. Curr Opin Neurobiol 10:732–739

Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299:1898–1902

Flaherty AW, Graybiel AM (1995) Motor and somatosensory corticostriatal projection magnifications in the squirrel monkey. J Neurophysiol 74:2638–2648

Fuster JM (1991) The prefrontal cortex and its relation to behavior. Prog Brain Res 87:201–211

Goldman-Rakic PS (1996) The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. Philos Trans R Soc Lond B Biol Sci 351:1445–1453

Graybiel AM (2008) Habits, rituals, and the evaluative brain. Annu Rev Neurosci 31:359–387

Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35:4–26

Haber SN, Kim KS, Mailly P, Calzavara R (2006) Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. J Neurosci 26:8368–8376

Hikosaka O, Takikawa Y, Kawagoe R (2000) Role of the basal ganglia in the control of purposive saccadic eye movements. Physiol Rev 80:953–978

Hikosaka O, Nakamura K, Nakahara H (2006) Basal ganglia orient eyes to reward. J Neurophysiol 95:567–584

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

Hori Y, Minamimoto T, Kimura M (2009) Neuronal encoding of reward value and direction of actions in the primate putamen. J Neurophysiol 102:3530–3543

Houk JC, Adams JL, Barto AG (1995) A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: Houk JC, Davis JL, Beiser DG (eds) Models of information processing in the basal ganglia. The MIT Press, Cambridge, pp 249–270

Inokawa H, Yamada H, Matsumoto N, Muranishi M, Kimura M (2010) Juxtacellular labeling of tonically active neurons and phasically active neurons in the rat striatum. Neuroscience 168:395–404

Ito M, Doya K (2009) Validation of decision-making models and analysis of decision variables in the rat basal ganglia. J Neurosci 29:9861–9874

Kato M, Kimura M (1992) Effects of reversible blockade of basal ganglia on a voluntary arm movement. J Neurophysiol 68:1516–1534

Kawagoe R, Takikawa Y, Hikosaka O (1998) Expectation of reward modulates cognitive signals in the basal ganglia. Nat Neurosci 1:411–416

Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF (2006) Optimal decision making and the anterior cingulate cortex. Nat Neurosci 9:940–947

Kimura M, Minamimoto T, Matsumoto N, Hori Y (2004) Monitoring and switching of cortico-basal ganglia loop functions by the thalamo-striatal system. Neurosci Res 48:355–360

Kojima T, Onoe H, Hikosaka K, Tsutsui K, Tsukada H, Watanabe M (2009) Default mode of brain activity demonstrated by positron emission tomography imaging in awake monkeys: higher rest-related than working memory-related activity in medial cortical areas. J Neurosci 29:14463–14471

Lau B, Glimcher PW (2008) Value representations in the primate caudate nucleus during matching behavior. Neuron 58:451–463

Lauwereyns J, Watanabe K, Coe B, Hikosaka O (2002) A neural correlate of response bias in monkey caudate nucleus. Nature 418:413–417

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

Matsumoto K, Suzuki W, Tanaka K (2003) Neuronal correlates of goal-based motor selection in the prefrontal cortex. Science 301:229–232

McCoy AN, Platt ML (2005) Expectations and outcomes: decision-making in the primate brain. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 191:201–211

McFarland NR, Haber SN (2000) Convergent inputs from thalamic motor nuclei and frontal cortical areas to the dorsal striatum in the primate. J Neurosci 20:3798–3813

Middleton FA, Strick PL (2000) Basal ganglia and cerebellar loops: motor and cognitive circuits. Brain Res Brain Res Rev 31:236–250

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

Monchi O, Petrides M, Petre V, Worsley K, Dagher A (2001) Wisconsin card sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. J Neurosci 21:7733–7741

Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H (2004) Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. Neuron 43:133–143

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

Nambu A (2008) Seven problems on the basal ganglia. Curr Opin Neurobiol 18:595–604

Nambu A, Kaneda K, Tokuno H, Takada M (2002) Organization of corticostriatal motor inputs in monkey putamen. J Neurophysiol 88:1830–1842

Nicola SM (2007) The nucleus accumbens as part of a basal ganglia action selection circuit. Psychopharmacology (Berl) 191:521–550

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

Pasquereau B, Nadjar A, Arkadir D, Bezard E, Goillandeau M, Bioulac B, Gross CE, Boraud T (2007) Shaping of motor responses by incentive values through the basal ganglia. J Neurosci 27:1176–1183

Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. Nature 400:233–238

Redgrave P, Prescott TJ, Gurney K (1999) Is the short-latency dopamine response too short to signal reward error? Trends Neurosci 22:146–151

Rudebeck PH, Behrens TE, Kennerley SW, Baxter MG, Buckley MJ, Walton ME, Rushworth MF (2008) Frontal cortex subregions play distinct roles in choices between actions and stimuli. J Neurosci 28:13775–13785

Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. Science 310:1337–1340

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

Sawaguchi T, Iba M (2001) Prefrontal cortical representation of visuospatial working memory in monkeys examined by local inactivation with muscimol. J Neurophysiol 86:2041–2053

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

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

Shidara M, Richmond BJ (2002) Anterior cingulate: single neuronal signals related to degree of reward expectancy. Science 296:1709–1711

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

Shima K, Tanji J (1998) Both supplementary and pre supplementary motor areas are crucial for the temporal organization of multiple movements. J Neurophysiol 80:3247–3260

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

Sutton RS, Barto AG (1998) Reinforcement learning. The MIT press, Cambridge

Takada M, Tokuno H, Hamada I, Inase M, Ito Y, Imanishi M, Hasegawa N, Akazawa T, Hatanaka N, Nambu A (2001) Organization of inputs from cingulate motor areas to basal ganglia in macaque monkey. Eur J Neurosci 14:1633–1650

Tricomi E, Balleine BW, O'Doherty JP (2009) A specific role for posterior dorsolateral striatum in human habit learning. Eur J Neurosci 29:2225–2232

Watanabe M, Cromwell HC, Tremblay L, Hollerman JR, Hikosaka K, Schultz W (2001) Behavioral reactions reflecting differential reward expectations in monkeys. Exp Brain Res 140:511–518

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

Yun RJ, Krystal JH, Mathalon DH (2010) Working memory overload: fronto-limbic interactions and effects on subsequent working memory function. Brain Imaging Behav 4:96–108