Role of midbrain dopamine neurons in reward-based learning

Takemasa Satoh

Division of Neurobiology, Department of Biomedical Sciences,
School of Life Science, Faculty of Medicine, Tottori University

Nishimachi 86, Yonago, Tottori 683-8503, Japan

January 2005

CONTENTS

This thesis is based on the following articles.

1. **Correlated coding of motivation and outcome of decision by dopamine neurons.**
   Takemasa Satoh, Sadamu Nakai, Tatuo Sato and Minoru Kimura.
   Journal of Neuroscience, 23: 9913-9923 (2003)

2. **Coding property to saliency by midbrain dopamine neurons during reward-based learning.**
   Takemasa Satoh, Sadamu Nakai, Tatuo Sato and Minoru Kimura.

General Introduction

Animals can adapt their behaviors to the environment through their experiences. The consequences of actions, which come as feedback from the environment, play an important role to learn and maintain appropriate behavior. Thorndike's Law of Effect states that an action followed by an appetitive consequence will occur more frequently, but an action followed by an aversive consequence will be less likely to occur (Thorndike, 1911). Appetitive consequences are referred to reward or positive reinforcer (Skinner, 1953). Such food, water and sex are examples of innate reward. The reinforcement effects of rewards are thought to be fundamental processes on the instrumental conditioning (Skinner, 1953; Schultz et al., 1997; Berridge, 2001), and based on the acquisition and maintenance of behavior chaining such as multiple sequential movements and verbal behavior (Skinner, 1953; Suri and Schultz, 1998; Malott and Suarez, 2003).

The midbrain dopamine containing (DA) neurons are the best candidate for the reward and reinforcement mechanisms. Depletion of dopamine in the dorsal striatum resulted in deficits not only in learning procedural strategies for sequential motor tasks (Matsumoto et al., 1999), but also in expressing learned responses of striatal neurons to conditioned stimuli (CS) (Aosaki et al., 1994). Schultz and colleagues revealed that DA neurons are strongly activated by delivery of an unpredictable reward, decrease discharges by omission of an expected reward, and maintain an baseline activity when fully expected reward is delivered (Schultz et al., 1993; Mirenowicz and Schultz, 1994). They hypothesized that DA neurons represent errors of reward prediction (REEs) and proposed that REEs play a role to general teaching signal during reward-related learning (Schultz et al., 1997; Hollerman and Schultz, 1998; Schultz, 1998).

On the other hand, for a voluntary action to occur, animals must be motivated to do it, in addition to know how to do it. Although it is clear that learning rates and task performances are influenced by motivational and drive states (Skinner, 1953), most of reinforcement theories of DA function mentioned little about the importance of motivation. Psychological theories suggested that motivational state has at least two fundamental effects on behavior learning and decision-making. First, motivation has incentive and evocative effect on the goal-directed and reward seeking behaviors (Lovibond, 1983; Berridge, 2001). For instance, motivational states modulate the ability of pavlovian incentive cue, which is associated with reward and elicit learned responses for reward (Cardinal et al., 2002). Second, motivational manipulations such as satiation and deprivation modulate the quality of reward (Berridge, 2001) and the effectiveness of reinforcer (Skinner, 1953; Michael, 1982, 2000).

Substantial amount of studies suggested that DA system is involved in motivational mechanisms (Wise et al., 1978; Robbins and Everitt, 1996; Koepp et al., 1998; Wyvell and Berridge, 2000). In addition, DA neurons also respond to biologically salient stimulus including novel, unexpected and intense sensory stimulus (Steinfels et al., 1983; Ljungberg et al., 1992; Horvitz et al., 1997; Horvitz, 2000). However how DA signals related to reward, motivation, attention, saliency and uncertainty (Fiorillo et al., 2003) are integrated and influence behavioral learning remains poorly understand.

In this thesis, I, in collaboration with Dr. Kimura, addressed this issue by introducing an instrumental conditioning task that monkeys made a series of behavioral decision based on trial-specific reward expectations. This task paradigm allowed us to examine how the activity of DA neurons represents reward expectation, its error and motivational properties. In addition, brief flash of visual cue (Flash) was used as salient stimulus to instruct the end of current block of trials and the start of new one. Thus this task paradigm allowed us to examine whether the activity of DA neurons related to behavioral switching of action selection.

Single neuron activity of DA neurons was recorded in two monkeys during this task. In chapter 1 of this thesis, I focused on the activity of DA neurons to CS and reinforcer. I found that half of DA neurons were responsive to CS. I investigated relationship among the magnitude of CS responses, degree of reward expectation and level of motivation to perform task trials. I have found for the first time that magnitude of responses to CS represents level of motivation to initiate task trials that was reflected in behavioral reaction times after CS. About 60% of DA neurons were responsive to reinforcer stimuli occurred after behavioral decisions as well as to CS. It was found that responses to reinforcer reflect positive and negative REEs quantitatively. To investigate whether the responses to CS and reinforcer act as general teaching signals for learning or as teaching signal for specific decisions among alternatives, I compared the magnitudes of the two types of responses at monkey's decisions to three target buttons separately. In addition, about a half of neurons were significantly responsive to both CS and reinforcer. To address functional roles of dual coding of motivational properties and REEs by single DA neurons, I investigated relationship between the magnitude of responses to CS and those to reinforcer in single trial type. I then examined the development of motivational properties and REEs signals through process of task learning.

In chapter2 of this thesis, I focused on the activity of DA neurons to salient stimulus during this task. I founded that small population of DA neurons responded to Flash. To investigate functional role of the responses of DA neurons to Flash, I analyzed relationship between behavioral switching of action selections and the responses to Flash. CS and Flash were same modality (visual) of sensory stimulus in this study. I then examined the difference of coding properties between CS and Flash.

References for General Introduction

Aosaki T, Graybiel AM, Kimura M (1994) Effect of the nigrostriatal dopamine system on acquired neural responses in the striatum of behaving monkeys. Science 265:412-415.

Berridge KC (2001) REWARD LEARNING: Reinforcement, Incentives, and Expectations. In: The Psychology of Learning and Motivation, Volume 40, pp 223-278. San Diego: Academic Press.

Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci Biobehav Rev 26:321-352.

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

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

Koepp MJ, Gunn RN, Lawrence AD, Cunningham VJ, Dagher A, Jones T, Brooks DJ, Bench CJ, Grasby PM (1998) Evidence for striatal dopamine release during a video game. Nature 393:266-268.

Lovibond PF (1983) Facilitation of instrumental behavior by a Pavlovian appetitive conditioned stimulus. J Exp Psychol Anim Behav Process 9:225-247.

Malott RW, Suarez EAT (2003) Principles of behavior. New York: Person Prentice Hall.

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

Michael J (1982) Distinguishing between discriminative and motivational functions of stimuli. J Exp Anal Behav 37:149-155.

Michael J (2000) Implications and refinements of the establishing operation concept. J Appl Behav Anal 33:401-410.

Mirenowicz J, Schultz W (1994) Importance of unpredictability for reward responses in primate dopamine neurons. J Neurophysiol 72:1024-1027.

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

Robbins TW, Everitt BJ (1996) Neurobehavioural mechanisms of reward and motivation. Curr Opin Neurobiol 6:228-236.

Schultz W (1998) Predictive reward signal of dopamine neurons. J Neurophysiol 80:1-27.

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

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

Skinner BF (1953) Science and human behavior. New York: Macmillan.

Spanagel R, Weiss F (1999) The dopamine hypothesis of reward: past and current status. Trends Neurosci 22:521-527.

Suri RE, Schultz W (1998) Learning of sequential movements by neural network model with dopamine-like reinforcement signal. Exp Brain Res 121:350-354.

Thorndike EL (1911) Animal intelligence. New York: The Macmillan Company.

Wise RA, Spindler J, deWit H, Gerberg GJ (1978) Neuroleptic-induced "anhedonia" in rats: pimozide blocks reward quality of food. Science 201:262-264.

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

Chapter 1:

Correlated coding of motivation and outcome of decision by dopamine neurons.

Summary

We recorded activity of midbrain dopamine neurons in an instrumental conditioning task in which monkeys made a series of behavioral decisions based on the distinct reward expectations. Dopamine neurons responded to the first visual cue appeared in each trial (conditioned stimulus, CS) through which monkeys initiated trial for decision while expecting trial-specific reward probability and volume. The magnitude of neuronal responses to CS was roughly proportional to reward expectations but with considerable discrepancy. On the other hand, the CS responses appear to represent motivational properties, because their magnitude at trials with identical reward expectation had significant negative correlation with reaction times of animal after the CS. Dopamine neurons responded also to reinforcers occurred after behavioral decisions, and the responses precisely encoded positive and negative reward expectation errors (REEs). The gain of coding REEs by spike frequency increased during learning act-outcome contingencies through a few months of task training, while coding of motivational properties remained consistent during the learning. We found that the magnitude of CS responses was positively correlated with that to reinforcers. This suggested a modulation of the effectiveness of REEs as a teaching signal by a motivation. For instance, rate of learning could be faster when animals are motivated, while slower when less motivated, even at identical REEs. Therefore, the dual, correlated coding of motivation and REEs suggested involvement of dopamine system both in reinforcement in more elaborate ways than currently proposed and in motivational function in reward-based decision-making and learning.

Introduction

Rewards such as food, sex and money are critically involved in the processes of decision-making (Herrnstein and Vaughn, 1980; Arnauld and Nichole, 1982) and behavioral learning (Thorndike, 1911; Hull, 1943; Rescorla and Wagner, 1972). The midbrain dopamine containing (DA) neurons are major neural substrate for the reward mechanisms.

Deprivation of dopamine in the dorsal striatum resulted in deficits not only in learning procedural strategies for performing sequential motor tasks (Matsumoto et al., 1999), but also in expressing learned responses of striate neurons to conditioned stimuli (CS) (Aosaki et al., 1994). Schultz and colleagues showed that DA neurons respond to reward during initial phase of learning, but respond to CS associated with reward in advanced stage of learning (Ljungberg et al., 1992; Schultz et al., 1993; Mirenowicz and Schultz, 1994). DA neurons increase discharges when a reward occurs unexpectedly, decrease discharges when expected reward is withheld and maintain a baseline discharge rate when reward is retrieved as expected (Schultz et al., 1997; Schultz, 1998). These observations led them to propose a hypothesis in which errors of reward expectation (REEs) are represented in the activity of DA neurons. Recently, they supported this hypothesis by finding that phasic responses of DA neurons varied monotonically with the change of reward probability (Fiorillo et al., 2003).

On the other hand, a substantial body of evidence suggests involvement of DA systems in the processes of motivation (Robbins and Everitt, 1996; Koepp et al., 1998; Salamone and Correa, 2002; Wise, 2002), in switching attentional and behavioral selections to salient stimuli that underlie associative learning (Redgrave et al., 1999; Spanagel and Weiss, 1999). It has been well documented that DA neurons show phasic activations by a wide variety of salient stimuli including novel and high intensity stimuli (Jacobs, 1986; Schultz and Romo, 1987; Ljungberg et al., 1992; Horvitz et al., 1997). Therefore, a critical question remains why DA neurons code several distinct signals, related to CS, reinforcement, uncertainty (Fiorillo et al., 2003), motivation and attention, and how these signals are integrated with the processes of decision-making and learning. This question has not been addressed.

In the present study, we investigated this issue specifically by examining the activity of DA neurons of monkeys that made a series of behavioral decisions based on trial-specific reward expectations. Neuronal responses to CS were roughly proportional to reward expectations but with considerable discrepancy. On the other hand, the CS responses appear

to represent motivational properties, because their magnitude at trials with identical reward expectation had significant negative correlation with reaction times of animal after the CS. The responses to reinforcer stimuli occurred after the behavioral decisions (outcomes) precisely encoded positive and negative REEs. The magnitude of responses to CS was positively correlated with that to outcome, suggesting modulation of REE coding by motivation. The dual, correlated coding of motivation and REEs suggested involvement of dopamine system not only in the reinforcement processes in more elaborate ways than currently proposed but also in motivational function in decision-making and learning.

Materials and Methods

Animals and surgery.

Two male Japanese monkeys (*Macaca fuscata*: monkey DN, monkey SK) were used in this study. 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 skulls using standard surgical procedures. The monkeys were sedated with ketamine hydrochloride (6 mg/kg, i.m.), and then anesthetized with sodium pentobarbital (Nembutal, 27.5 mg/kg, i.p.). Supplemental Nembutal (10 mg/kg/2 hrs, i.m.) was given as needed. The recording chamber was positioned at an angle of 45° in order to record the activity of dopamine neurons in the right midbrain under stereotaxic guidance.

Behavioral paradigm.

The monkeys were trained to sit in a primate chair facing a small panel that was placed 27 cm in front of their faces. On the panel were a small, rectangular push button with red light emitting diode (LED) (start LED, 14 x 14 mm) at the bottom, three push buttons with green LEDs (target LEDs, 14 x 14 mm) in the middle row, and a small, red LED (GO LED, 4 mm diameter) just above the center push buttons (Figure 1A). The task was initiated by illumination of the start LED on the push button. The monkeys depressed the illuminated start button with their left hand. The start LED was turned off 400 ms after the monkeys had continued to hold the button. Then, the target LEDs and a GO LED were simultaneously turned on. The monkeys were required to continue depressing the start button for variable lengths of time between 0.6 and 0.8 s before the GO LED was turned off. They released the start button and depressed one of the three illuminated target buttons. If an incorrect button was depressed, a beep sound with a low-tone (300 Hz for 100 ms) occurred with a delay of 500 ms, and the next trial began by illuminating the start LED at 7.5 s after releasing the depressed button. Because the monkey remembered the incorrect button selected at the first trial, it made a choice between the two remaining buttons. If the monkey made an incorrect choice again, the third trial started after a low-tone beep and the monkey depressed the remaining, single correct button. If the correct button was depressed, a beep sound with a

high-tone (1 k Hz for 100 ms) occurred with a delay of 500 ms, and a small amount of reward water was delivered through the spout attached to the monkey's mouth.

The high-tone and low-tone beep sounds served as positive and negative reinforcers, respectively, after the behavioral decisions. Once the monkeys found the correct button, the same button was used as the correct button in the succeeding trials. Thus, the monkeys received a reward three times by selecting the same button during three consecutive trials. Two seconds after releasing the depressed button, the three target buttons were flashed at the same time for 100 ms to inform the animal of the end of a block of trials. At 3.5 s after the flashing of target buttons the next block of trials began with the correct button in a new, unpredictable location. Thus, the trials in a single block were divided into two epochs (Figure 1B). The first epoch was the trial-and-error epoch where the monkey searched for the correct button on a trial and error basis. Three types of trials occurred: trials in which the monkeys selected the correct button at the first, second or third choice in a single block (N1, N2 and N3, respectively). The second epoch was the repetition epoch in which the monkeys selected the known correct button in two successive trials after they had found the correct button during the trial-and-error epoch. Two types of trials occurred: the first and the second trials at the repetition epoch (R1 and R2, respectively). The amount of reward water was 0.35 ml in the trial-and-error epoch, and 0.2 ml in the repetition epoch.

Over the 7 months (monkey DN) and 3 months (monkey SK) of recording sessions in this study, there were substantial changes in the probabilities of correct button presses and rewards in each trial type, especially during the first month when the monkeys had not reliably acquired the trial type-specific expectation of reward. The average correct choice rates in the 5 trial types in the early, partially learned stage and the later, fully learned stage are summarized in Table 1. After the early, partially learned stage, we set the average correct choice rate at N1 to be lower than a chance level of 33.3%, and the actual average rate was 20.0 ± 8.6% in monkey DN and 16.8 ± 3.4% in monkey SK.

Data recording and analysis.

Single neuron activity was recorded using epoxy-coated tungsten microelectrodes (26-10-2L, Frederic Haer, Bowdoinham) with an exposed tip of 15 μm and impedances of 2-5 MΩ (at 1 kHz). The neuronal activity recorded by the microelectrodes was amplified and displayed on an oscilloscope using conventional electrophysiological techniques. Band-pass

filters (50 Hz to 3 kHz band-pass with a 6 dB/octave rolloff) were used. The action potentials of the 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 0.04 ms. 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 occurring during behavioral tasks. The electrodes were inserted through the implanted recording chambers and advanced by means of an oil-drive micromanipulator (MO-95, Narishige, Tokyo). We searched for dopamine neurons in and around the pars compacta of the substantia nigra (SNc). Electrode penetrations at an angle of 45° through the posterior putamen, the external and internal globus pallidus, and internal capsule before reaching the midbrain considerably assisted our approaches to the dopamine neurons because we have much experience in recording from the putamen and globus pallidus. In accordance with previous studies on the discharge properties of dopamine neurons (Grace and Bunney, 1983; Schultz, 1986), we identified dopamine neurons based on the following four criteria. First, the action potentials of the dopamine neurons have a relatively long duration [1.6 - 2.9 ms (range), 2.2 ± 0.3 ms (mean ± SD), Figure 3A]. Second, the background discharge rate of the dopamine neurons is low (0.5 - 7.4 impulses/s, 4.0 ± 1.6 impulses/s), and in sharp contrast to the high background discharge rate of neurons in the SNr. Third, under the histological reconstruction of electrode tracks in relation to electrolytic lesion marks (a total of 6 marks made by passing positive DC current of 25 μA for 30 sec), the recording sites were located in the SNc or VTA in monkey DN. Fourth, unexpectedly delivered reward water caused a phasic increase in the discharge rate.

Electromyographic (EMG) activity was recorded in the triceps and biceps brachii muscles (prime movers for the button press), and the digastric muscle (prime mover for consuming liquid rewards) of monkey DN through chronically implanted, multi-stranded, teflon-coated, stainless steel wire electrodes (AS631, Cooner Wire, California) with leads that led subcutaneously to the head implant. The EMG signals were amplified, rectified, integrated, and monitored on-line on a computer display along with the recorded neuronal activity. In small number of experiments (10 recording days in monkey DN, 5 days in monkey SK), eye movements were also monitored by measuring the corneal reflection of an infrared light beam through a video camera at a sampling rate of 250 Hz. A computer system (RMS R-21C-A, Iseyo-Denshi, Tokyo) determined the two-dimensional (x and y) signal of the center of

gravity of the reflected infrared light beam. The spatial resolution of this system was ca. ± 0.15°. The muscle activity and eye position signals from the video system were also fed to the laboratory computer through the A-D converter interface at a sampling rate of 100 Hz.

Distinct levels of reward expectation (REs, %) and reward expectation error (REEs, %) in the 5 types of trials (N1, N2, N3, R1 and R2) were estimated as:

REs (%) = probability of reward × 100

or

REs (%) = probability of reward × volume of reward (ml) × 100

Positive REEs (%) = (occurrence of reward (1) – probability of reward) × 100

or

Positive REEs (%) = (occurrence of reward (1) – probability of reward) × volume of reward (ml) × 100

Negative REEs (%) = (occurrence of no reward (0) – probability of reward) × 100

or

Negative REEs (%) = (occurrence of no reward (0) – probability of reward) × volume of reward (ml) × 100

The responses of neurons were determined in peri-event time histograms of the neuronal impulse discharges as an increase or decrease in the discharge rate after a behavioral event, relative to the discharge rate for 1000 ms preceding the presentations of the start LED and BEEP. The onset of a response was determined as the time point at which the change in the discharge rate achieved a significance level of $P < 0.05$ by the two-tailed Wilcoxon test (Kimura, 1986).

Histological examination.

After recording was completed, the monkey DN was anesthetized with an overdose of pentobarbital sodium (90 mg/kg, i.m.) and transcardially perfused with cold heparinized 0.9% NaCl solution followed by 4% paraformaldehyde in 0.1M phosphate buffer. Frozen sections were cut at every 50 μm at planes parallel to the recording electrode penetrations. The

sections were stained with cresyl violet. We reconstructed the electrode tracks and recording sites of the DA neurons based on the 6 electrolytic microlesions (Figure 3B,C). Sections spaced at 300-μm intervals through the striatum and substantia nigra were stained for tyrosine hydroxylase-like immunoreactivity (Chemicon anti-TH, 1:1000, Matsumoto et al., 1999).

Results

Evolution of task performance through learning act-outcome relations

Monkeys chose one of three potentially correct buttons as their first choice (Fig. 1*A* and *B*, trial and error epoch, see Experimental Procedures). They got 0.35 ml of reward water if the choice was correct. However, if it was incorrect, they made a second choice from the remaining two buttons. If the choice was incorrect again, they chose the remaining one button and received a reward in all of the trials except for about 10 % of trials in which monkeys made errors again. Thus, there were 3 types of trials -- N1, N2 and N3 -- in which the monkeys hit the correct button as their first, second or third choice, respectively, on a trial and error basis. Once the monkeys found the correct button, they obtained a smaller amount (0.2 ml) of reward water by choosing the previously found correct button for 2 succeeding trials (Fig. 1*B*, repetition epoch, R1 and R2 trials). Figure 1*C* plots the average correct choice rates in the 5 trial types over the entire recording period in monkey DN.

Through performing the task for 7 months in monkey DN and 3 months in monkey SK, the monkeys learned the rules for the "choice-among-3" task after having learned the rules for "choice-between-2" task. The task performance of the monkeys changed during the course of learning. At the initial stage of learning, the average rate of correct choices for the N1 trials was around 1/3, as theoretically predicted. At the later stage, the correct choice rate in the N1 trials was controlled to be 20% so that the monkeys would expect a reward at much lower probability than in the N2 (1/2) and N3 trials (1/1). In this study, it was critical that monkeys had wide range of trial-specific reward expectations before making behavioral decisions. The average correct choice rates in the 5 trial types at the initial, partially learned stage, and at the later, fully learned stage are shown in Fig. 2*A* and are summarized in Table 1. The variation in the daily average correct choice rate for each trial type became much smaller in the fully learned stage than in the early stage. There was a clear tendency to choose the button that had been a correct one in the previous set of trials during the N1 trials in both monkey DN (average, 62%) and monkey SK (96%).

The briskness of depressing the start button after the appearance of the start LED, the first behavioral reaction at each trial, also changed during learning in a trial type-dependent manner. The start LED acted as conditioned stimulus (CS) with respect to the unconditioned

stimulus (US, reward) in the present instrumental conditioning task. The average reaction times (RTs) of button pressing after presentation of a CS in all 5 trial types were relatively prolonged during the initial stage of learning (days 1 to 38, 533.3 ± 49.7 ms in monkey DN; day 1 to 15, 486.6 ± 31.1 ms in monkey SK, mean ± SD). There was no significant difference among the RTs in the 5 trial types in monkey DN (one-way ANOVA, $F_{4,35}$=1.176, P>0.3). In monkey SK, there was also no significant difference among the RTs in the 5 trial types except between R1 and N1 (one-way ANOVA, $F_{4,90}$=4.353, P=0.038, post hoc Scheffe test). After this stage, the RTs of N2, N3, R1 and R2 trials became much shorter (450.2 ± 38.0 ms in monkey DN; 468.0 ± 32.7 ms in monkey SK), while those of the N1 trials became longer (559.0 ± 52.9 ms in monkey DN; 531.2 ± 32.0 ms in monkey SK). The RTs in the N1 trials were significantly longer than those in the 4 other trial types ($F_{4,215}$=62.744, P<0.0001 in monkey DN; $F_{4,180}$=28.682, P<0.0001 in monkey SK, post hoc Scheffe test). Based upon these learning stage-dependent differences in task performance, the experimental sessions were separated into an early stage, a partially learned stage, and later, fully learned stage. The average RTs at the two stages in the two monkeys, thus defined, are plotted in Fig. 2*C* and *D*.

The monkey DN developed a characteristic orofacial reaction after incorrect trials at the fully learned stage. The electromyograms (EMGs) of digastric muscle activity during the consumption of the liquid reward revealed similar activity patterns for all five types of correct trials (Fig. 2*E*). During the incorrect trials, by contrast, the digastric muscle was much more strongly activated because of the characteristic orofacial reaction. Interestingly, the reaction occurred in a trial type-specific manner. Large activation occurred in the N2 and N3 trials, with the maximum activation in the N3 trials in which reward probability was highest in the trial and error epoch (Fig. 2*E*), while the activity in the N1 trials in which the reward probability was lowest was smallest, though slightly larger than that in the correct trials. Probably, the muscle activation reflects levels of animal's disappointment at incorrect choices with no reward, because disappointment would be greater when reward expectation was higher. EMGs at R1 and R2 trials are not shown because of very small number of incorrect trials in these trial types. These observations indicated that the monkeys gradually developed both an understanding of reward probabilities and volumes -- thus the expectation of reward -- and the levels of motivation specific to each trial type through learning act-outcome relations in the present reward-based decision-making task.

Identification of midbrain DA neurons

In two monkeys, we recorded the activity of 253 presumed DA neurons (163 in monkey DN and 90 in monkey SK) in the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA) while the monkeys made a series of reward-based decisions. These neurons had characteristic discharge properties that have been used to identify DA neurons (Grace and Bunney, 1983; Schultz, 1986), such as the long duration of the action potential and tonic discharges at approximately 4 impulses/s (see Experimental Procedures). The properties of these DA neuron discharges significantly differed from those of neurons in the nearby substantia nigra pars reticulata (SNr), as shown in Fig. 3*A*. In addition, an unexpectedly delivered water reward caused a phasic increase in the discharge rate of most of the DA neurons examined (20/25 in monkey DN).

In this study, we describe the activity of 52 DA neurons from monkey DN and 56 DA neurons from monkey SK that maintained consistent discharge rates and responsiveness during more than 50 correct trials in the task for at least 30 min. In monkey DN, the locations of these 52 neurons were histologically verified in the midbrain (Fig. 3*B* and *C*). DA neurons recorded in the SNc and VTA of monkey DN will be described as a single population in this study. In monkey SK, neuronal recording is still in progress and, thus, histological examination has not been made. However, the characteristic depth profiles of neuronal activity through oblique microelectrode penetrations were very similar in the two monkeys. For instance, there were abrupt shifts from low background discharges in the putamen to very high background discharges with thin action potentials in the globus pallidus. The electrode entered the internal capsule with low neural noise, then, entered into either the area with slow tonic discharges of thick action potentials characteristic of the SNc and VTA or the area with very high frequency discharges of thin spikes characteristic of the SNr. Therefore, the activity of 56 presumed DA neurons identified on this basis is described as a separate neuronal population for monkey SK from that for monkey DN.

Responses to conditioned stimulus (CS) in the instrumental conditioning

The DA neurons increased or decreased their tonic discharges after two different sensory events occurred in the task. One was the CS that instructed the monkeys to initiate each trial of the instrumental task. The second one was a high-tone or low-tone beep sound reporting

either that the animal's choices were correct and that the reward would come or that the choices were incorrect and no reward would be given. The high-tone and low-tone beep sounds after the animal's choices thus acted as positive and negative reinforcers. The rest of the events -- such as GO LED, hand movements and reward -- did not evoke significant modulations of DA neuron activity.

The DA neurons produced a brisk response to the CS (Fig. 4*A*). The magnitude of the responses varied trial by trial. It was found that the variation of response magnitude occurred in a trial type-dependent manner ($P>0.1$ in monkey DN; $P<0.001$ in monkey SK, one way ANOVA), as shown in the responses of a single neuron and ensemble average responses in Fig. 4*A* and *B*. In Fig. 4*C* and *D* are plotted against trial type average increases or decreases of discharges from the baseline level in response to CS in two monkeys. About a half of neurons showed significant responses to CS (Table 2, 27/52 neurons in monkey DN; 27/56 neurons in monkey SK). The results are based on the neuronal activity both in the partially learned and fully learned stages. The average response was an increase in the discharges in all of the trial types. In monkey SK, the responses in N1 trials were the smallest among the 5 trial types ($P<0.05$, post hoc Fisher's PLSD) and the responses in the N3 trials were larger than those in the other trial types ($P<0.05$, post hoc Fischer's PLSD).

What is the functional significance of the trial type-dependent responses of DA neurons to CS? The responses may represent animal's expectation of reward, because it is supposed in the reinforcement learning algorithm that the responses to CS represent weighted sum of predicted future reward, the value function (Sutton, 1988; Sutton and Barto, 1998). We tested this hypothesis by comparing the response magnitudes with the reward expectation. Reward expectations at each trial type could be estimated in this study in terms of either the probability of reward or the product of probability and volume of reward (Fig. 4*C* and *D*). The neural responses and the reward expectations are normalized to have the same value at the trial type with maximum reward expectation (N3 in the case of product of probability and volume, R1 or R2 in the case of probability). The curve of the reward expectations as the probability of reward (open squares) did not predict the DA neuron responses in both monkeys, although the responses were smallest consistently at N1 trials in which reward expectation was lowest among the 5 trial types. The reward expectations as the product of probability and volume of reward (filled circles) did not estimate the responses very well too, although they explained a decrease of responses at R1 and R2 trials.

We tested an alternative hypothesis that the responses to a CS may reflect animal's motivation to work for a reward. We used a time for monkeys to depress the start button (reaction times, RTs) after it (CS) was presented as an index of how much the monkeys were motivated to work at the trial, because RTs are one of behavioral measures reflecting levels of motivation (Konorski, 1967; Shidara et al., 1998; Watanabe et al., 2001; Kobayashi et al., 2002; Takikawa et al., 2002). To dissociate an involvement of reward expectation in the CS responses from that of motivation, we studied the correlation of RT and amplitude of DA neuron response within single trial types in which monkeys performed the trial with consistent level of reward expectations. Figure 5A shows ensemble averages of neuronal activity of the 3 groups of R2 trials with short, middle and long RTs in monkey SK. Largest activation occurred in short RT trial group, smallest activation occurred at longest RT group and middle level of activation occurred in the middle RT group. Fig. 5*B* and *C* plots the average magnitude of CS responses against the RTs in each trial type. There was a significant negative correlation between the neuronal responses and the RTs in both monkey DN (e.g., N1, $r = -0.277$, $p<0.001$) and monkey SK (N1, $r = -0.252$, $p<0.01$). But within the same groups of RT, there was no significant difference among the CS responses at different trial types ($P>0.05$, one way ANOVA) except for RT 500-600 ms group in monkey SK ($P<0.01$). The negative correlation was also observed on a single trial basis (Fig. 5*D*, $r= -0.191$, $p<0.001$, N1 trials in monkey DN), and on a single neuron basis (Fig. 5*E*, $r= -0.224$, 56 neurons in monkey SK). The single trial-based negative correlation in at least one trial type was observed in 9 out of 52 neurons in monkey DN and 12 out of 56 neurons in monkey SK.

In most of N1 trials, monkeys chose the button that had been a correct one in the previous set of trials (average 52% in monkey DN; 98% in monkey SK). There was no significant influence of the tendency on the neuronal responses to CS at N1 trials ($P>0.3$ for monkey DN, Wilcoxon rank test). Measurement of eye position signals during task performance revealed that monkeys were looking at either one of three target buttons or a hold button before illumination of the hold button (CS) in most of time. Specifically, during 500 ms before the CS appearance, monkeys tended to look at hold button more often at R1 and R2 trials than at N1 trials. Thus, the difference in eye positioning before the CS could be related to variance in RT of depressing hold button. But limited amount of eye movement data in the present study did not allow us to draw definitive conclusions on this issue. It is our important future issue.

To study the origin of the large variations in the RTs within a single trial type, trials were classified into those performed early in the session (initial 3 hours) of the daily experimental schedule when the monkeys were thirsty, and those performed during the later session (after the initial 3 hours) when the monkeys became less thirsty or experienced satiety after receiving a certain amount of reward water. The RTs in each trial type during the early session were shorter than those during later session by 21 ± 5.3 ms (mean ± SD in 5 trial types, $P<0.001$ in N1, N2, N3 and R2, Mann-Whitney U test), in monkey DN, and by 10.6 ± 12.5 ms in monkey SK ($P<0.05$ in N2 and N3, Mann-Whitney U test). Consistent changes in the RTs in the weekly schedule of experiments were also observed. In monkey SK, RTs were longer on Monday than those on the other days of the week by 25.9 ± 22.7 ms (mean ± SD in 5 trial types, $P<0.01$ for N2 and N3, Mann-Whitney U test). This was probably because the monkeys spent weekends with free access to food and water, and were less motivated to work on Monday. Thus, these results support the motivation hypothesis.

Fig. 6*A* shows the population response histograms of 3 groups of N2 trials with 3 behavioral choices (LEFT, CENTER and RIGHT) to target buttons in monkey SK. There was no clear difference among the ensemble activities of 3 groups. Fig. 6*B* and *C* plots the average responses to CS in all 5 trial types. Although magnitudes of responses to CS varied dependent on the trial type, there was no significant difference among the responses to CS at the same trial type but different succeeding choices of target button ($p>0.27$ in monkey DN; $p>0.29$ in monkey SK, Kruskal-Wallis test).

Coding outcomes of behavioral decision

DA neurons characteristically responded to the reinforcers after behavioral decisions. Figure 7*A* illustrates representative responses of a DA neuron in the SNc of monkey DN to the positive reinforcer after correct choices (all 5 trial types) and to the negative reinforcer after incorrect choices (N1 and N2). The neuronal responses to the positive reinforcer consistently produced an increase in the discharges rate, positive response. In contrast, negative reinforcer produced the decrease of discharges, negative response that was preceded in many cases by small transient increase of discharges. The population response histograms of the 52 neurons in monkey DN in Fig. 7*B* demonstrated a systematic dependency of neuronal responses on the trial type. These relations were consistently observed in the

ensemble activity of 52 neurons in monkey DN and 56 neurons in monkey SK. The magnitudes of average positive and negative responses are plotted in Fig. 7*C* and *D*. In both monkeys, the positive responses were the highest at N1, became smaller at N2, and became still smaller at N3 trials. There was nearly no response at the repetition epoch (R1 and R2 trials) in monkey DN (Fig. 7*C*), while small responses in monkey SK (Fig. 7*D*). More than 60% of neurons responded to the reinforcers (Table 2, 32/52 in monkey DN; 39/56 in monkey SK). The recording sites of neurons responsive only to start cue, those responsive only to reinforcers and those responsive to both start cue and reinforcers in monkey DN were histologically reconstructed in the midbrain (Fig. 9). But it did not appear to be a special tendency of distribution of the three classes of neurons in the midbrain. One-way ANOVA revealed that the trial type had a significant effect on the positive DA neuron responses ($F_{4,255}$=28.425, P<0.0001 in monkey DN; $F_{4,275}$=13.594, P<0.0001 in monkey SK, post hoc Scheffe test). Although not statistically significant ($F_{3,174}$=0.564, P>0.6 in monkey DN; $F_{2,151}$=1.332, P>0.2 in monkey SK, one-way ANOVA), the negative responses to the negative reinforcer also changed in a trial type-dependent manner.

What is the functional significance of the systematic dependencies of both positive and negative responses toward the trial type? We assessed the claim that the responses represent reward expectation errors (REEs). Positive and negative REEs derived from product of probability and volume of reward at each trial type (see Experimental Procedures) are superimposed on the response histograms in Fig. 7*C* and *D*. The two plots are normalized so that the same value at the maximum REEs occurred in N1 for positive responses and in R1 or R2 for negative responses. It was found that the magnitudes of the positive responses for each trial type could be estimated surprisingly well by the REEs. The positive responses were significantly correlated with the REEs in both monkey DN (r=0.627; p<0.001) and monkey SK (r=0.399; p<0.001). The gain of coding REEs was 0.083 impulses/% REEs in monkey DN and 0.026 impulses/% REEs in monkey SK. There was a weaker correlation between the negative responses and the REEs in monkey DN (r=0.087; gain, 0.007 impulses/% REEs) and in monkey SK (r=0.159; p<0.05; gain, 0.009 impulses/% REEs).

To examine a possibility that responses to reinforcer modulated by behavioral decision to target buttons, ensemble and single neuronal activities of DA neurons to reinforcers are separated with behavioral selections to target buttuons. Fig. 8*A* shows the population response histograms of 3 groups of N2 trials with 3 behavioral choices (LEFT, CENTER and RIGHT)

to target buttons in monkey DN. There was no clear difference among the ensemble activities of 3 groups. Fig. 8*B* and *C* plots the average responses to the reinforcers in 5 correct and 2 incorrect (N1 and N2) trial types. There was no significant difference among the average reinforcer responses at the same trial type but different choice of target button in both monkeys ($p>0.3$ in monkey DN; $p>0.07$ in monkey SK, Kruskal-Wallis test). At single neuron level, 2 out of 52 neurons in monkey DN and 4 out of 56 neurons in monkey SK show significant differences ($p<0.05$, Kruskal-Wallis test).

The positive responses at N1 trials after monkey DN chose the button that had been a correct one in the previous set of trials (average 7.11 ± 0.98 impulses/s) were slightly larger than those when previously incorrect button was chosen (average 5.46 ± 0.90 impulses/s, $P<0.05$, Wilcoxon rank test). This could reflect the difference in either the level of motivation or reward expectation between the two groups of N1 trials. There was no significant difference in negative responses ($P>0.7$, Wilcoxon rank test).

In summary, these observations indicate that the responses of DA neurons to positive reinforcer after behavioral decisions precisely encode REEs. The responses to negative reinforcer also encode REEs, although the gain in encoding by decreasing the discharge rate is smaller than that for the positive reinforcer. In addition, coding of REEs was not modulated by which target button monkeys chose. In other words, there was no significant difference in magnitude of DA neuron responses at different button selection, if the trial type, consequently REEs, was the same.

Positive relation between the responses to CS and those to outcomes of decision

What kind of roles does the simultaneous coding of motivational properties and REEs by single DA neuron activity play? To address this issue, we studied the relation of responses to CS and those to positive reinforcers (high tone beep) in single trial type. Because the responses to CS in N1 trials and those to positive reinforcers in R1 and R2 trials were very small and because the number of N3 trials was very small, the responses in N2 trials were quantitatively examined. Responses to CS were positively correlated with those to positive reinforcers in monkey DN (Fig. 10*A*, $r=0.234$, 52 neurons) and in monkey SK (Fig. 10*B*, $r=0.524$, 56 neurons, $p<0.001$). The positive correlation was also observed in N1 trials in monkey SK but not in monkey DN. The results, thus, support an interesting view that the number of DA neuron spikes encoding REEs, gain of coding REEs, might be positively

modulated by the responses to CS that appear to reflect levels of motivation.

Development of coding reward-related information during learning

In parallel with the evolution of each animal's task performance during the initial and late stages of learning (Fig. 2), DA neurons modified their response properties during the two learning stages. Figure 11*A* plots responses to CS as a function of RTs in two learning stages. In two monkeys, the CS responses were negatively correlated with RTs in both partially learned and fully learned stages. More interestingly, slope of the correlation was consistently maintained in the two stages of learning in two monkeys, although the correlation in the partially learned stage of monkey DN was not significant probably because small number of neurons were studied (n=8). In the initial stage of learning the "choice-among-3" task but after having learned the "choice-between-2" task, DA neurons did not show robust responses to reinforcer stimuli occurred after animal's behavioral decisions in both monkeys. Remarkably, responses were so small and variable in monkey DN that average responses at incorrect N1 and N2 trials were not negative but positive (Fig. 11*B*, left panel). By contrast, in the fully learned stage when RTs after start cue at N1 trials became significantly longer than those at the other 4 trial types because of very low reward expectation, much stronger positive responses appeared in correct N1 and N2 trials (Fig. 11*B*, right panel). About 4-fold increase occurred in the gain of coding positive REEs in monkey DN. Similar, but mild, increase was also observed in monkey SK. On the other hand, there was no apparent change in the gain of coding negative REEs through learning. This was in sharp contrast to the responses to the CS in which the slopes of negative correlation between the responses to CS and RTs was consistently maintained through learning (Fig. 11*A*).

These observations indicated that the coding of REEs by DA neuron activity develops through the process of learning act-outcome relations in the reward-based decision-making task, while motivational properties attributed to CS appear at an initial stage of learning and are maintained during learning.

Table 1. Correct Choice Rates in Each Trial Type at the Early and Late Stages of Learning

Correct choice rates in monkey DN					
Learning stage	N1	N2	N3	R1	R2
Early stage* (8 DA cells)	29.6 ± 15.6	49.7 ± 13.2	75.2 ± 14.1	98.4 ± 2.3	99.5 ± 1.5
Late stage** (44 DA cells)	18.3 ± 5.3	51.5 ± 10.0	89.9 ± 8.6	99.0 ± 2.1	98.3 ± 2.4
Early + Late stages (52 DA cells)	20.0 ± 8.6	51.2 ± 10.4	87.7 ± 10.8	98.9 ± 2.1	98.5 ± 2.3
Correct choice rates in monkey SK					
Learning stage	N1	N2	N3	R1	R2
Early stage*** (19 DA cells)	17.2 ± 4.6	48.1 ± 10.9	76.2 ± 10.1	97.8 ± 4.0	93.9 ± 5.3
Late stage**** (37 DA cells)	16.8 ± 3.4	48.9 ± 8.4	92.1 ± 8.6	99.1 ± 1.4	99.3 ± 2.0
Early + Late stages (56 DA cells)	16.9 ± 3.8	48.7 ± 9.2	86.7 ± 11.8	98.7 ± 2.6	97.4 ± 4.3

Results are expressed as the means ± standard deviation of percentages. * Early stage of learning, day 1 - day 36 of the study. **Late stage of learning, day 37 - month 7. ***Early stage of learning, day 1 - day 15 of the study. **** Late stage of learning day 16 - month 3.

Table 2. Number of Responsive DA Neurons to Conditioned Stimulus (CS) and to Reinforcers

Monkey DN		
	Partially Learned Stage	Fully Learned Stage
Conditioned Stimulus (CS)	4	23
Reinforcers	2	30
CS and Reinforcers	1	18
Total	8	44

Monkey SK		
	Partially Learned Stage	Fully Learned Stage
Conditioned Stimulus (CS)	11	16
Reinforcers	12	26
CS and Reinforcers	7	16
Total	19	37

Figures are number of responsive neurons determined by Wilcoxon single rank test at $p<0.05$ (Kimura, 1986).

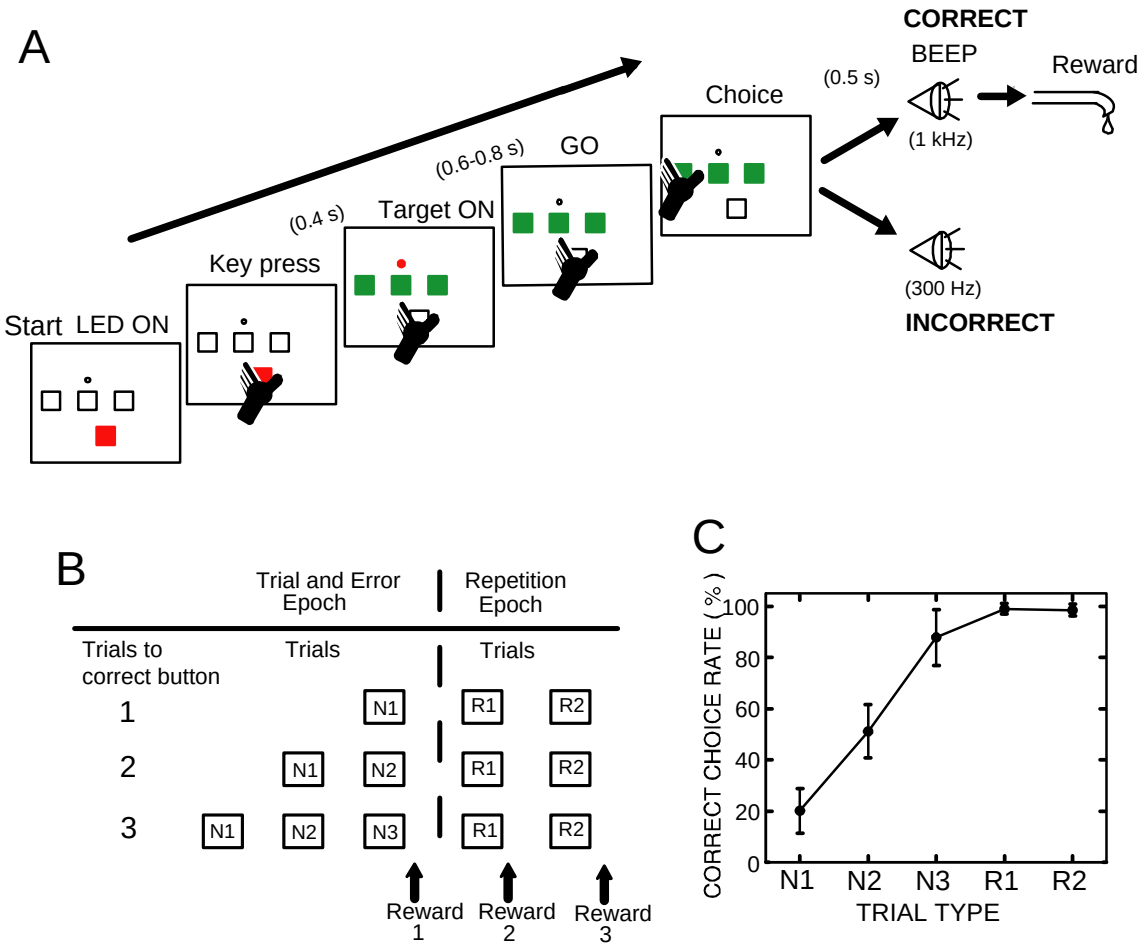

Figure 1. Behavioral task, trial types and percent correct at each trial type *A*, Illustration of sensorimotor events that appeared during a single trial. See details in the Experimental Procedures. *B*, Two epochs (trial-and-error epoch and repetition epoch) and 5 trial types (N1, N2, N3, R1, R2) in a block of trials classified on the bases of correct and incorrect button choices. *C*, Correct choice rate over the 7-month study as a function of trial type in monkey DN. The results are expressed as means and SD of all trials during which all DA neuron activity was recorded.

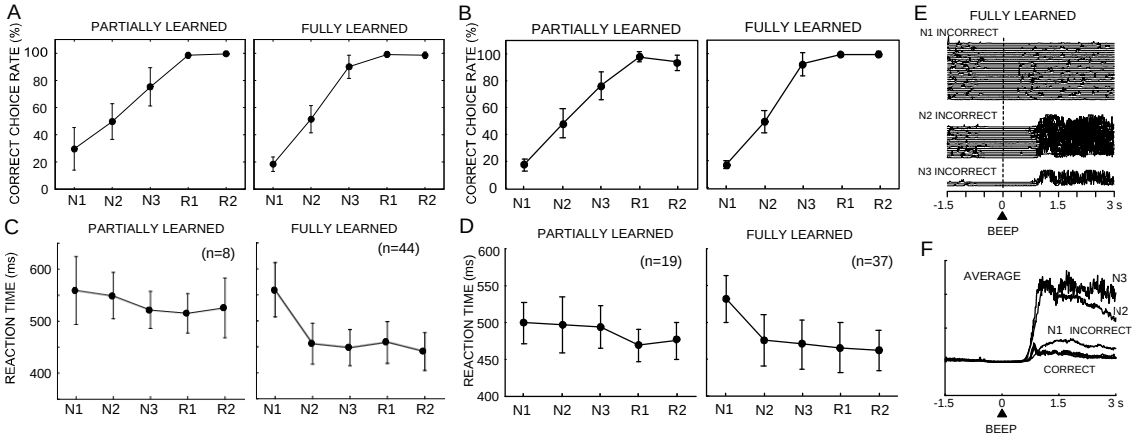

Figure 2. Task performance in the partially learned and fully learned stages *A*, Correct choice rate against the trial types in the partially learned (first 1-36 days) and fully learned stages (37-215 days) in monkey DN. *B*, Same as (*A*), but in the partially learned (first 1-15 days) and fully learned stages (16-95 days) for monkey SK. *C*, Average RTs for the start LED at each trial type in monkey DN. Error bars indicate SD. *D*, Same as (C) but for monkey SK. *E*, Superimposed traces of orofacial muscle activity during 3 incorrect trial types (N1, N2, N3) (left), and average traces during 5 correct (N1, N2, N3, R1, R2) and 3 incorrect (N1, N2, N3) trial types (right) in monkey DN. BEEP indicates the onset of the beep sound after animal's choices.

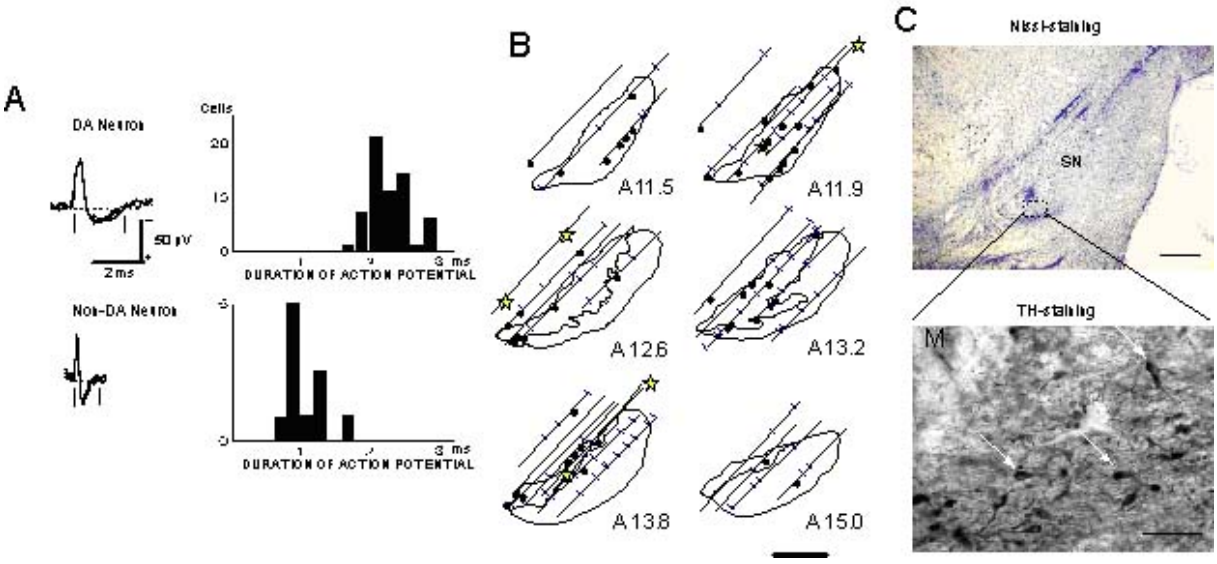

Figure 3. Electrophysiological and histological identification of DA neurons *A*, On the left are superimposed traces of extracellularly recorded action potentials of DA (SNc) and non-DA neurons (SNr). The two vertical lines and the horizontal interrupted line indicate how the duration of the action potential was measured. Histograms of the duration of recorded action potentials are shown on the right panel. *B*, Histological reconstruction of the recording sites of DA neurons (filled circles) and non-DA neurons (blue lines) along electrode tracks in and around the SNc. Stars indicate locations of electrolytic micro-lesion marks. Scale: 2 mm. *C*, A Nissl-stained section at the level of the substantia nigra (SN) is shown (scale, 1 mm) (top), and part (interrupted circle) of the neighboring, TH-stained section is shown at higher magnification (scale, 100 µm) (bottom). White arrows indicate TH-immunoreactive neurons. M indicates part of a lesion mark.

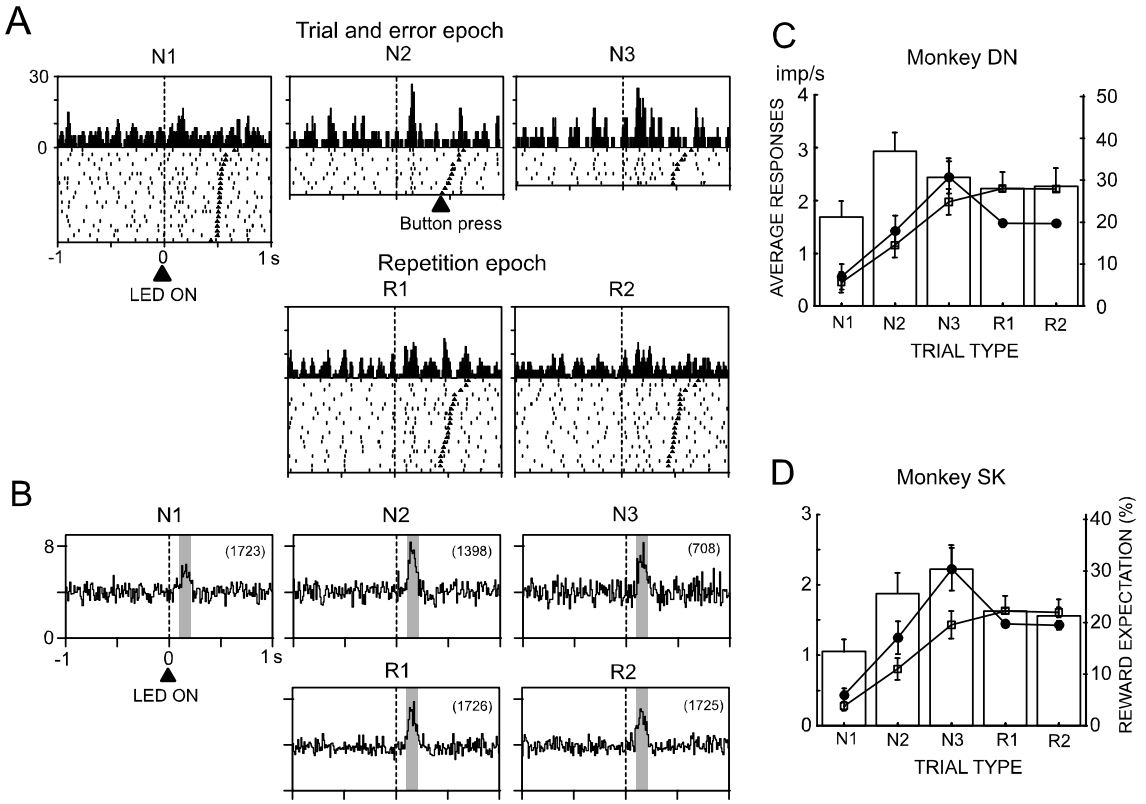

Figure 4. Response of DA neurons to the start LED (CS) *A*, Activity of a single DA neuron recorded in the SNc of monkey DN before and after CS in the 5 trial types. Impulse discharges that occurred during the individual trial types are represented separately as rasters and histograms. The activity is centered at the onset of CS (vertical interrupted line). The trials in the raster display were reordered based on the time interval between onset of CS and depression of the start button. The time point of the button press in each trial is marked on the raster. *B*, Population response histograms of 52 DA neurons to CS in monkey DN. *C*, Average increase in the discharge rate of the 52 DA neurons during the fixed time window indicated by the shaded areas in each histogram in (*B*), relative to the discharge rate over the 500-ms period just preceding the onset of CS. The results are shown as means ± S.E. in monkey DN. On the response histogram are superimposed curves of reward expectations, as a probability (open squares) and a product of probability and volume of reward (filled circles, explanation, see text). Scale of the reward expectation on the ordinate on the right side is for the product of probability and volume of reward. *D*, Same as (*C*) but for 56 DA neurons in monkey SK. The bin width of the histograms was 15 ms.

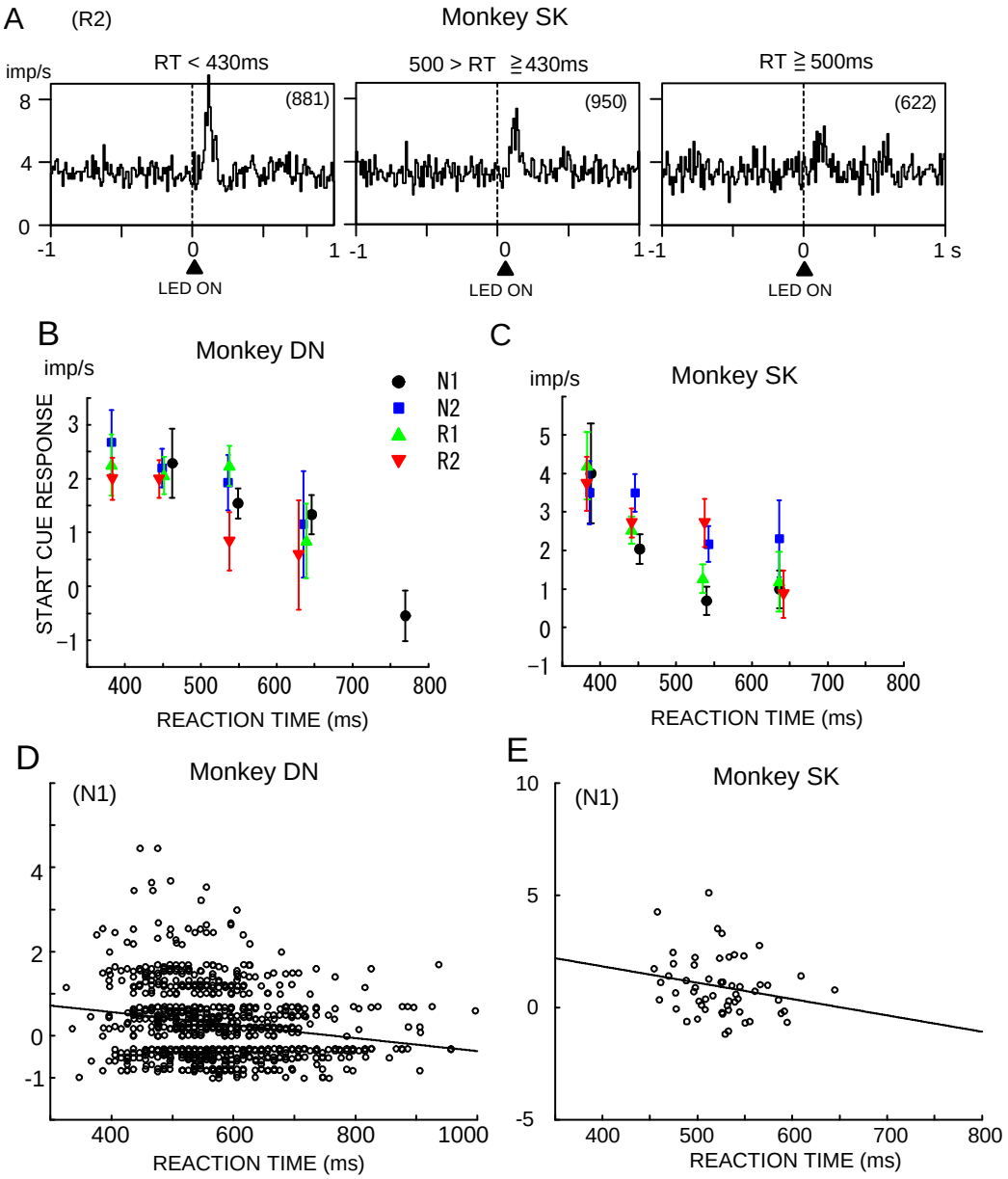

Figure 5. Relation of response magnitudes of DA neurons to briskness of behavioral responses to CS *A*, Population response histograms of the 56 DA neurons in monkey SK to CS during the R2 trials. Histograms are separated based upon the trials with short, middle and long RTs to CS. The figures in parenthesis are the number of trials involved in each histogram. *B*, Correlation of magnitude of neural responses to CS in 52 DA neurons in monkey DN to RTs to depress the start button after CS. The correlations are plotted separately in N1, N2, R1 and R2 trials. The results of N3 trials are not plotted because of very small number of trials. The trials were classified into 5 groups based on the RTs, and the mean and SEM of DA neuron responses in these groups of trials are plotted. *C*, Same as (*B*) but for monkey SK based on the RTs of trials during recording of 56 DA neurons. Because the RTs in monkey SK were shorter than those in monkey DN by about 80 ms on average, the ranges of RTs in 3 groups of trials in monkey SK were shifted to shorter RTs from those in monkey DN. *D*, Correlation of magnitude of responses to CS with RTs in each trial in monkey DN. The correlation analysis was performed on 854 trials from 27 neurons showing significant

responses to CS. *E*, Correlation between average CS responses of single neurons and average RTs in monkey SK (56 neurons).

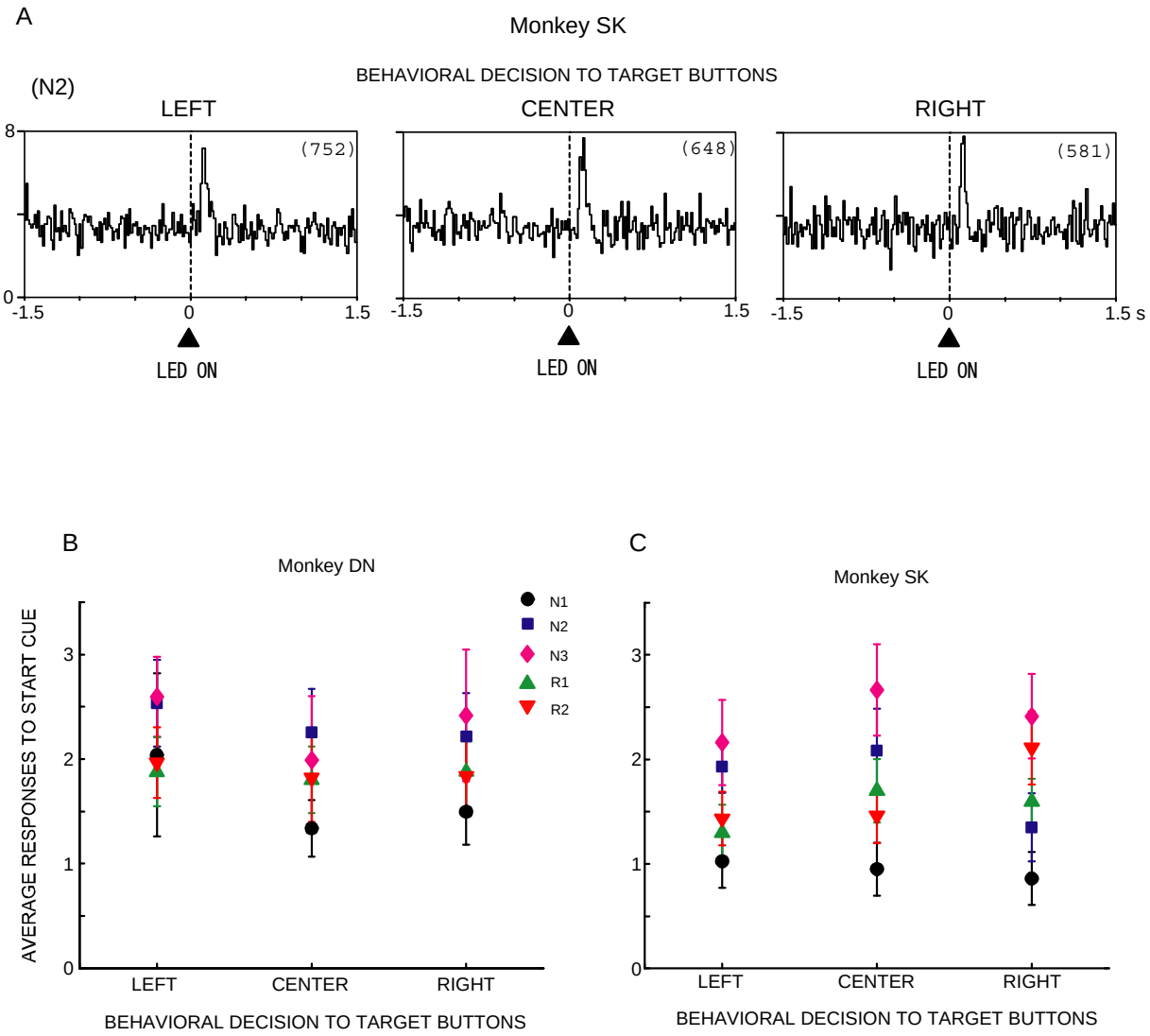

Figure 6. Relationship between response of DA neurons to CS and animal's choice to target buttons *A*, Population response histograms of 56 DA neurons in monkey SK to CS during N2 trials. Histograms were separated based upon trials with behavioral decision to 3 target buttons (LEFT, CENTER, RIGHT). The figure in parentheses indicates the number of trials used to obtain the population response. Bin width = 15 ms. *B*, Scatterplot of the average responses to CS (mean and SEM) and animal's choice to target buttons in monkey DN. The values are plotted separately in 5 trial types (N1, N2, N3, R1 and R2). *C*, Same as (*B*), but for monkey SK.

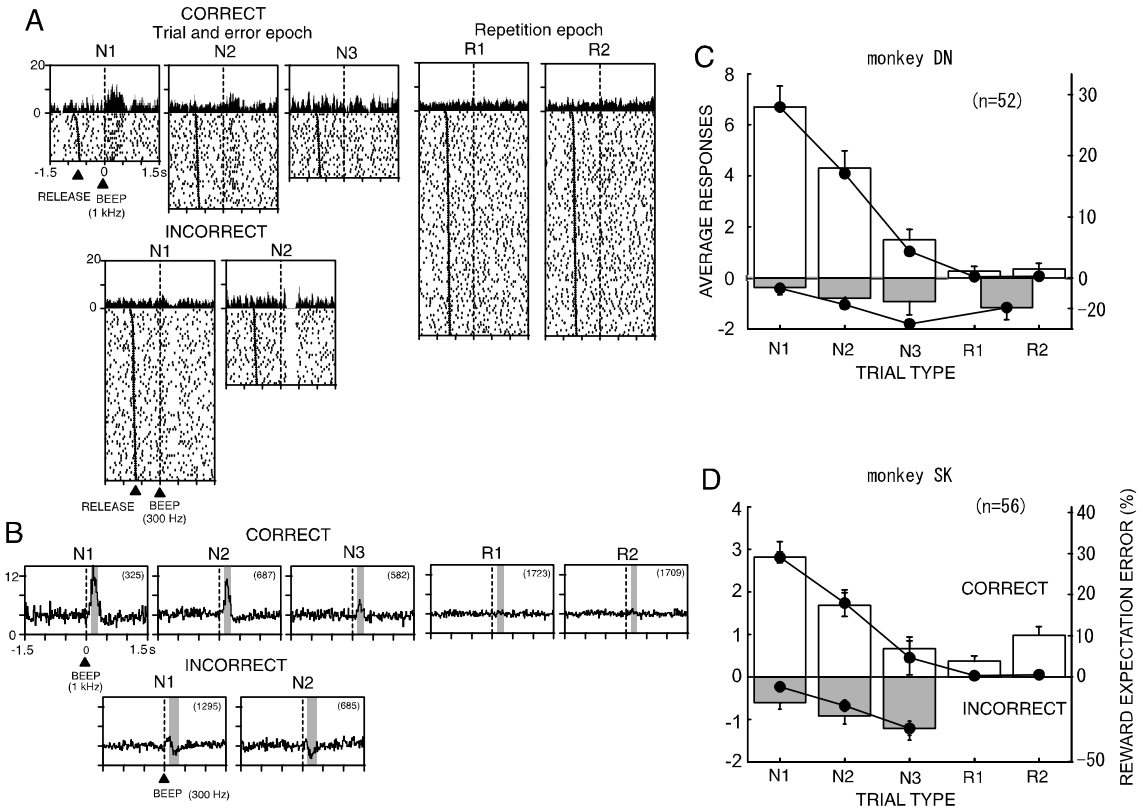

Figure 7. Responses of DA neurons to reinforcers after the animal's choices at each task trial *A*, Activity of a representative DA neuron at correct and incorrect choices in the 5 trial types. The displays are centered at the onset of the reinforcers (vertical interrupted lines). The trials in the raster display were reordered according to the time interval between the GO signal and onset of the reinforcers, and the time point of the GO signal in each trial is marked on the raster display. RELEASE indicates the time point at which the monkey released the start button to depress one of the target buttons. *B*, Population response histograms of 52 DA neurons in monkey DN during correct and incorrect choices in the 5 trial types. The figure in parentheses indicates the number of trials used to obtain the population response. *C*, The histogram of responses in monkey DN. The responses are shown as mean and SEM (vertical bar above or below each column) of the increase (correct trials) or decrease (incorrect trials) in the discharge rate during fixed time windows indicated by the shaded area in each histogram in (*B*), relative to the discharge rate during the 500-ms period just preceding the onset of CS. On the response histogram are superimposed positive and negative REEs (filled circles) derived from product of probability and volume of reward at each trial type (see Experimental Procedures). *D*, Same as (*C*) but for monkey SK. Because incorrect trials rarely occurred during the repetition epoch, the neuronal responses and REEs for the R1 and R2 trials were either combined and plotted as a single trial type in monkey DN or not shown in monkey SK.

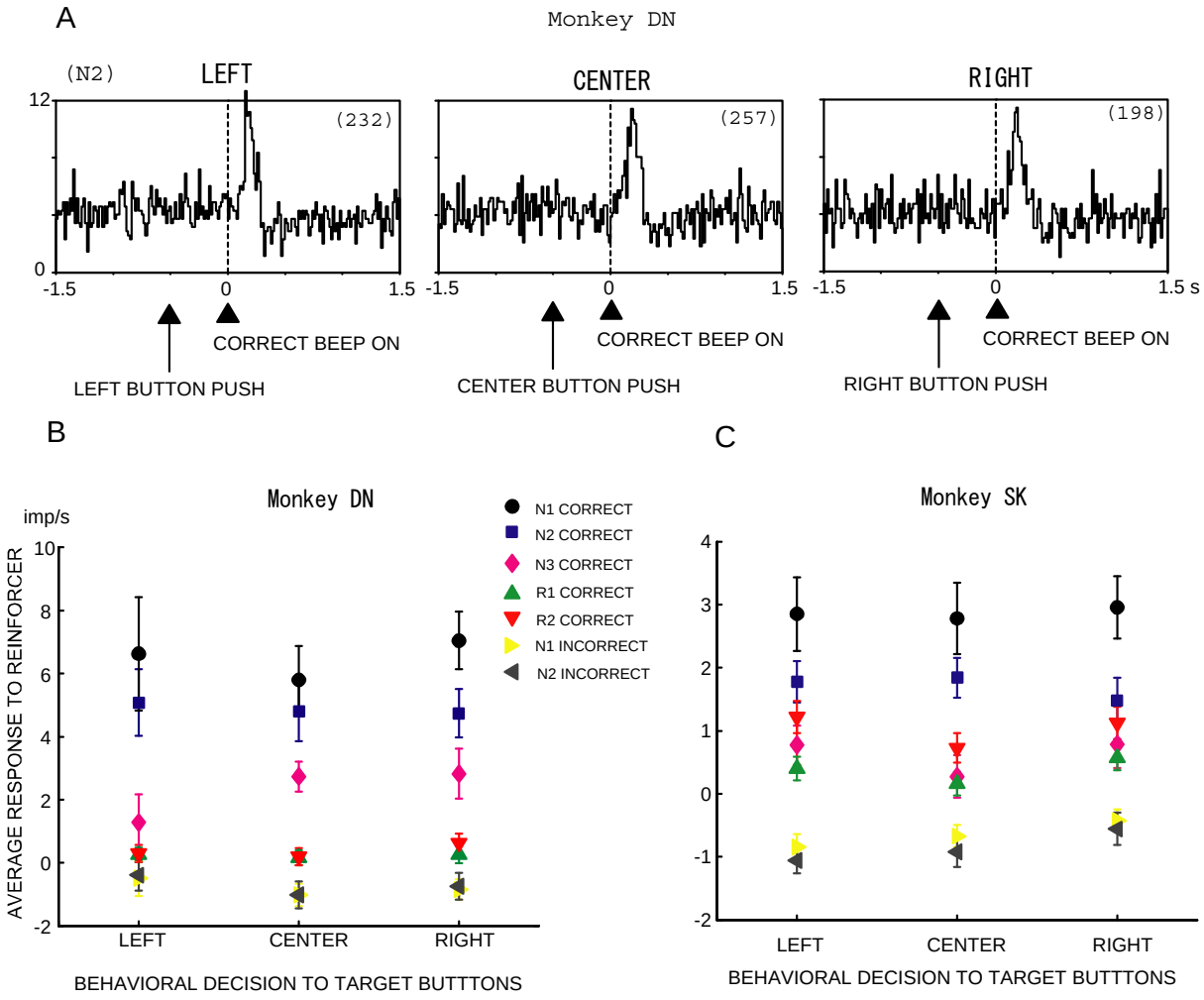

Figure 8. Relationship between the response magnitudes to reinforcers and behavioral decision to target buttons A, Population histograms of responses of 52 DA neurons in monkey DN to reinforcers during N2 trials. Histograms are separated based upon the trials with 3 animal's choices to 3 target buttons (LEFT, CENTER and RIGHT). *B*, Scatterplot of the average responses to reinforcers (mean and SEM) and animal's choice to target buttons in monkey DN. The values are plotted separately in 5 correct (N1, N2, N3, R1 and R2) and 2 incorrect (N1, N2) trial types. *C*, Same as (*B*), but for monkey SK.

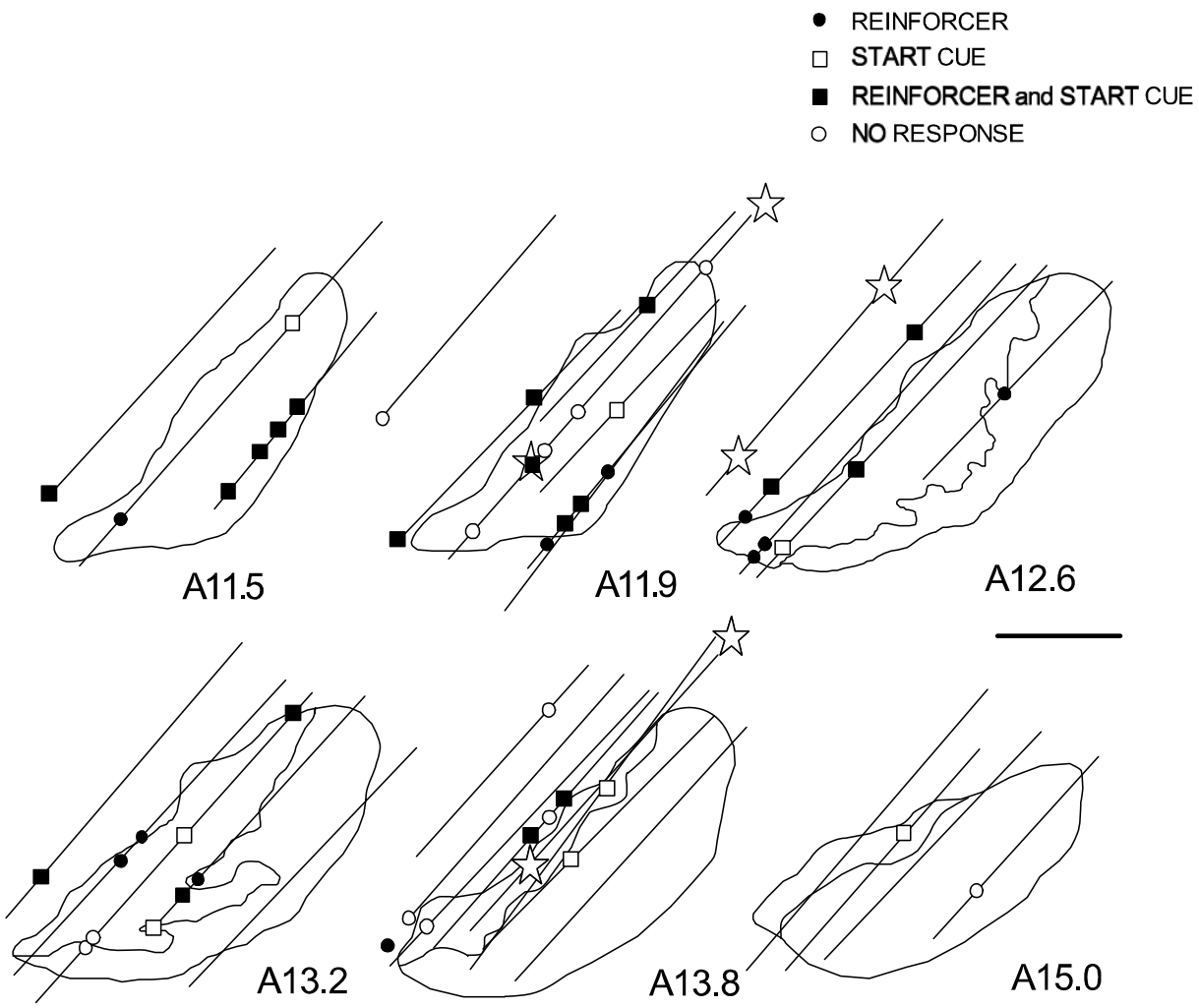

Figure 9. Positions of dopamine neurons are labeled according to the 3 task relationships or no responsive activity (Open Circle) along to electrode tracks. Filled circles indicate positions of neurons responsive to only reinforcer, open squares indicate those responsive only to start cue and filled squares indicate those responsive to both start cue and reinforcer. Scale: 2mm.

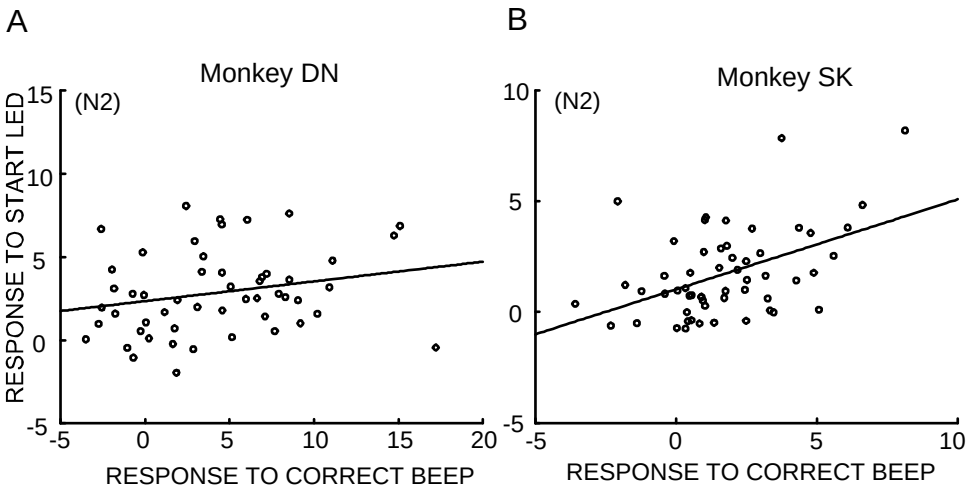

Figure 10. Relation of responses to CS and response to high tone beep, positive reinforcer *A*, Scatter plot showing positive correlation between the response to CS and response to positive reinforcer in N2 trials in monkey DN (r=0.234, slope=0.125). *B*, Same as (A) but for monkey SK (r=0.524, slope=0.551).

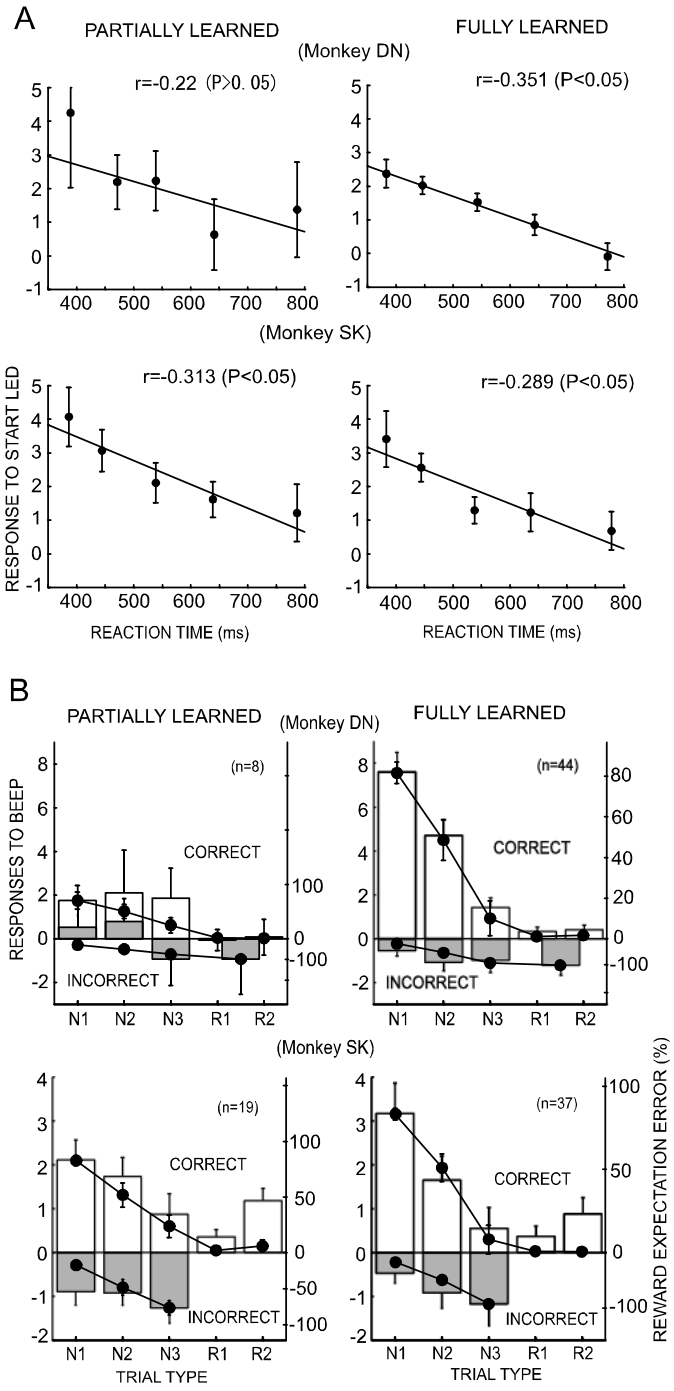

Figure 11. Responses of DA neurons at the early, partially learned stage and later, fully learned stage *A*, Scatter plot of the average responses of DA neurons (mean and SEM) and RTs to depress start button after the CS. The plots were made for all of the trials independent of trial type. Trials were divided into 5 groups based on the RTs. Regression lines are superimposed. *B*, Histograms of the responses of the DA neurons to the reinforcers after the animal's choices in the partially learned stage and fully learned stage in the 5 trial types. The values in the incorrect R1 and R2 trials are combined in monkey DN and are not plotted in monkey SK because of the very small number of trials. REEs (mean and SEM) are

superimposed on the histograms. The response histograms and REEs are normalized so as to have the same value at the maximum REE.

Discussion

The present study revealed, for the first time, three aspects regarding the properties of DA neuron activity. First, the responses of DA neurons to CS appear to represent motivational properties attributed to CS. Second, the responses of DA neurons to positive reinforcer after behavioral decisions precisely encode REEs. The responses to negative reinforcer also encode REEs, but the gain of encoding by decreasing discharge rate is much smaller than that for positive reinforcer. This finding is in agreement with those of Schultz and colleagues that phasic responses of DA neurons varied monotonically across the full range of reward probabilities (Fiorillo et al., 2003). But, we demonstrated directly, for the first time, that the DA neuron activity represents REEs in a quantitative manner in a series of reward expectation-based decision processes in the instrumental conditioning paradigm. In addition, we found that the responses to CS were positively correlated with those to reinforcers encoding REEs, suggesting modulation of efficacy of teaching signals by motivational process. Third, the precise coding of REEs by DA neuron activity develops through learning of act-outcome relations through a remarkable increase of gain of coding, while coding motivational attribution to CS appears at an initial stage of learning and is consistently maintained through the entire learning process.

Dual and correlated coding of motivation and of reward expectation error

The importance of reward in learning and decision-making has long been emphasized along two theoretical lines. First, reinforcement theories assume that reward learning consists primarily of a process by which behavior is directly strengthened or weakened by the consequence that follows it (Thorndike, 1911). Reinforcement learning theories proposed a computational algorithm of reward learning in which the agent adapts its behaviors based on errors of reward prediction as a teaching signal (Sutton, 1988; Sutton and Barto, 1998). Second, Pavlovian incentive theories suggest that if a stimulus becomes associated with primary reward, not only the Pavlovian association between the stimulus and a conditioned response occurs, but also a motivational transformation occurs. That is, the stimulus takes on specific motivational, incentive, properties (CS) that were originally possessed only by the primary reward itself (Bolles, 1972; Bindra, 1978; Toates, 1986; Dickinson and Balleine, 1994). The findings that the responses of DA neurons to reinforcers after an animal's choices

precisely encode REEs provide a solid experimental support to the models of reinforcement learning. These models suggest that DA neurons transmit REEs as reinforcement signals derived from a sum of the reward predictions at successive times that act like temporal derivatives (TD) and reward received to the target in the dorsal and ventral striatum and frontal cortices (Sutton, 1988; Barto, 1995; Houk et al., 1995; Montague et al., 1996; Schultz et al., 1997; Suri and Schultz, 1998; Sutton and Barto, 1998). Coding of positive REEs by increase of DA neuron spikes, thus increase of dopamine release, would facilitate adaptive changes of synaptic transmission related to reward-based learning in the target structures. Negative REEs were also coded by decrease of DA neuron spike rate, although gain of coding was low probably because of the floor effects in which decrease in the discharge rate saturates. Therefore, it is possible that the coding negative REEs by DA neurons might contribute to extinction or unlearning of actions, like the case of teaching signals (climbing fiber activity) in the cerebellar learning (Medina et al., 2002). Encoding positive and negative REEs might also suggest alternative functions to the reinforcement, such as the expectation of having to switch to a new behavioral strategy (acquire reward) or stick with the old one (wait for the next start signal). But the fact that the magnitude of responses to high and low tone beeps was precisely estimated by REEs appears to favor reinforcement over the switching between two strategies.

The responses to CS were not accurately estimated by the expectation of reward as a reward probability. Reward expectation as a product of probability and volume of reward better predicted responses to CS, while having considerable discrepancy still. On the other hand, we found, for the first time, that the responses to CS were correlated significantly with RTs at trials with identical level of reward expectation. Thus, it appears that the responses to CS might reflect participation in the processes of motivation, while apparently representing reward expectation. This could be the reason why the magnitude of responses to CS was not accurately estimated by the expectation of reward.

Interestingly, a similar negative correlation was recently reported between the magnitude of positive responses of the pedunculopontine tegmental nucleus (PPTN) neurons to the fixation point onset (CS) for an eye movement task and the reaction time of eye fixation after the CS (Kobayashi et al., 2002). In addition, the magnitude of PPTN neuron responses was positively correlated with the correct performance rate. These results suggested that the PPTN system might be involved in the processes of motivational and attentional control of

movement and in the neuronal mechanisms for reinforcement learning (Dormont et al., 1998; Brown et al., 1999; Kobayashi et al., 2002). Monosynaptic axonal projections from the pedunculopontine tegmental nucleus (PPTN) to DA neurons in SNc have been demonstrated (Futami et al., 1995). Thus, the PPTN is a strong candidate for a brain structure that supplies midbrain DA neurons with Pavlovian incentive-related signals.

The present study revealed that about a half of DA neurons studied significantly responded to both CS and reinforcers (Table 2, 19/52 in monkey DN; 23/56 in monkey SK). What is the functional significance of the dual coding of incentive attribution to CS and of REEs in reward-based decision-making and learning? One possible and fascinating role is a modulation of the effectiveness of REEs as teaching signal by a motivation. For instance, rate of learning could be faster when animals are highly motivated because of stronger activation of DA neurons, and thus larger amount of DA release, while slower when less motivated, even at identical REEs as a consequence of an action. Actually in the present study, there was a positive correlation between the responses to CS and those to positive reinforcers (Figure 7). This suggests a new and richer model for DA neurons as teaching signals in reinforcement learning than currently proposed. On the other hand, it is consistent with the theory of classical conditioning in which rate of learning is assumed to be under the influence of factors, such as attention or motivation (Rescorla and Wagner, 1972; Dickinson, 1980). In a computational point of view, an involvement of motivational process in the instrumental conditioning was recently emphasized, and a new model of reinforcement learning was put forward in which DA neurons transmit both reward expectation error and impact of motivation (Dayan and Balleine, 2002; McClure et al., 2003).

The principal functions of reward are supposed to produce satisfaction, to elicit approach behaviors, and to reinforce immediately preceding actions (Thorndike, 1911; Hull, 1943; Olds and Milner, 1954). The DA neuron responses to CS may participate in producing satisfaction and eliciting approach behaviors, while those to reinforcers may play a major part in the reinforcement. A special emphasis has been put on the reinforcement function for the DA neuron responses to the reinforcers, while relatively little attention have been paid to the responses to CS and to their functional significance. On the other hand, it has been well documented that DA neurons show phasic activations by a wide variety of salient stimuli including novel and high intensity stimuli (Jacobs, 1986; Schultz and Romo, 1987; Ljungberg et al., 1992; Horvitz et al., 1997). The responses to CS observed in the present study probably

share in their properties with the previously reported responses to salient stimuli. Animals would approach or escape from those stimuli that gain reinforcing efficacy by means of their association with appetitive or aversive stimuli, conditioned reinforcers. This process must play fundamental but distinct roles for behavioral decisions and learning from reinforcement process. This view was supported by the observation that the responses to CS and to reinforcer behaved independently during learning. The gain of coding the motivation estimated by RTs did not change during learning, suggesting the invariance of the mechanism linking the incentive attribution to CS with DA release in the dorsal and ventral striatum and frontal cortices during learning. In contrast, there was a remarkable elevation of the gain for encoding REEs by DA neuron spike density, thus DA release during learning. A critical question arises. Why were the target stimuli or GO signals ineffective to elicit DA neuron responses? Analysis of eye movements revealed a tendency of monkeys to make saccade before and after the CS frequently to one of targets which was going to be chosen after GO signal, and suggested that the monkeys made a decision at around the CS with trial type-dependent reward expectations. These observations could support the view that the DA responses to CS 'motivate' the whole trial. Understanding of this issue will be an important direction for our future research.

Although, in the present study, we emphasize the motivational properties for DA neuron activity, it is possible that the process of attention allocated to CS is also involved, because attention can contribute to shaping new forms of behaviors toward the direction of their goal, i.e., the reward (see also Dayan et al., 2000; Horvitz, 2000) and is difficult to estimate in separation from the motivation.

References

Aosaki T, Graybiel AM, Kimura M (1994) Effect of the nigrostriatal dopamine system on acquired neural responses in the striatum of behaving monkeys. Science 265:412-415.

Arnauld A, Nichole P (1982) The Art of Thinking: Port-Royal Logic. Indianapolis: Bobbs-Merrill.

Barto AG (1995) Adaptive critics and the basal ganglia. In: Models of Information Processing in the Basal Ganglia (Beiser DG, ed), pp 215-232. Cambridge: The MIT Press.

Bindra D (1978) How adaptive behavior is produced: a perceptual-motivation alternative to response reinforcement. Behavioral and Brain Sciences 1:41-91.

Bolles RC (1972) Reinforcement, expectancy, and learning. Psychological Review 79:394-409.

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

Dayan P, Balleine BW (2002) Reward, motivation, and reinforcement learning. Neuron 36:285-298.

Dayan P, Kakade S, Montague PR (2000) Learning and selective attention. Nat Neurosci 3 Suppl:1218-1223.

Dickinson A (1980) Contemporaly Animal Learning Theory. Cambridge: Cambridge Univ. Press.

Dickinson A, Balleine B (1994) Motivational control of goal-directed action. Animal Learning and Behavior 22:1-18.

Dormont JF, Conde H, Farin D (1998) The role of the pedunculopontine tegmental nucleus in relation to conditioned motor performance in the cat. I. Context-dependent and reinforcement-related single unit activity. Exp Brain Res 121:401-410.

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

Futami T, Takakusaki K, Kitai ST (1995) Glutamatergic and cholinergic inputs from the pedunculopontine tegmental nucleus to dopamine neurons in the substantia nigra pars compacta. Neurosci Res 21:331-342.

Grace AA, Bunney BS (1983) Intracellular and extracellular electrophysiology of nigral dopaminergic neurons--1. Identification and characterization. Neuroscience 10:301-315.

Herrnstein RJ, Vaughn WJ (1980) The allocation of individual behavior. In: Limits to Action (Staddon JER, ed), pp 143-176. New York: Academic.

Horvitz JC (2000) Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. Neuroscience 96:651-656.

Horvitz JC, Stewart T, Jacobs BL (1997) Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. Brain Res 759:251-258.

Houk JC, Adams JL, Barto AG (1995) A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: Models of Information Processing in the Basal Ganglia (Beiser DG, ed), pp 249-270. Cambridge: The MIT Press.

Hull CL (1943) Principles of behavior, an introduction to behavior theory. New York: D. Appleton-Century Co.

Jacobs BL (1986) Single unit activity of brain monoamine-containing neurons in freely moving animals. Ann N Y Acad Sci 473:70-77.

Kimura M (1986) The role of primate putamen neurons in the association of sensory stimuli with movement. Neurosci Res 3:436-443.

Kobayashi Y, Inoue Y, Yamamoto M, Isa T, Aizawa H (2002) Contribution of pedunculopontine tegmental nucleus neurons to performance of visually guided saccade tasks in monkeys. J Neurophysiol 88:715-731.

Koepp MJ, Gunn RN, Lawrence AD, Cunningham VJ, Dagher A, Jones T, Brooks DJ, Bench CJ, Grasby PM (1998) Evidence for striatal dopamine release during a video game. Nature 393:266-268.

Konorski J (1967) Integrative activity of the brain. Chicago: University of Chicago Press.

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

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

McClure SM, Daw ND, Montague PR (2003) A computational substrate for incentive salience. Trends Neurosci 26:423-428.

Medina JF, Nores WL, Mauk MD (2002) Inhibition of climbing fibres is a signal for the extinction of conditioned eyelid responses. Nature 416:330-333.

Mirenowicz J, Schultz W (1994) Importance of unpredictability for reward responses in primate dopamine neurons. J Neurophysiol 72:1024-1027.

Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16:1936-1947.

Olds J, Milner P (1954) Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. J Comp Physiol Psychol 47:419-427.

Redgrave P, Prescott TJ, Gurney K (1999) Is the short-latency dopamine response too short to

signal reward error? Trends Neurosci 22:146-151.
Rescorla RA, Wagner AR (1972) Current research and theory. In: Classical Conditioning II (Prokasy WF, ed), pp 64-99. New York: Appleton Century Crofts.
Robbins TW, Everitt BJ (1996) Neurobehavioural mechanisms of reward and motivation. Curr Opin Neurobiol 6:228-236.
Salamone JD, Correa M (2002) Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. Behav Brain Res 137:3-25.
Schultz W (1986) Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. J Neurophysiol 56:1439-1461.
Schultz W (1998) Predictive reward signal of dopamine neurons. J Neurophysiol 80:1-27.
Schultz W, Romo R (1987) Responses of nigrostriatal dopamine neurons to high-intensity somatosensory stimulation in the anesthetized monkey. J Neurophysiol 57:201-217.
Schultz W, Apicella P, Ljungberg T (1993) Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. J Neurosci 13:900-913.
Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593-1599.
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.
Spanagel R, Weiss F (1999) The dopamine hypothesis of reward: past and current status. Trends Neurosci 22:521-527.
Suri RE, Schultz W (1998) Learning of sequential movements by neural network model with dopamine-like reinforcement signal. Exp Brain Res 121:350-354.
Sutton RS (1988) Learning to predict by the method of temporal differences. Mach Learn 3:9-44.
Sutton RS, Barto AG (1998) Reinforcement Learning. Cambridge: The MIT press.
Takikawa Y, Kawagoe R, Itoh H, Nakahara H, Hikosaka O (2002) Modulation of saccadic eye movements by predicted reward outcome. Exp Brain Res 142:284-291.
Thorndike EL (1911) Animal intelligence. New York: The Macmillan Company.
Toates F (1986) Motivational systems. Cambridge, UK: Cambridge University Press.
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.
Wise RA (2002) Brain reward circuitry: insights from unsensed incentives. Neuron 36:229-240.

Chapter 2:

Coding properties of saliency by midbrain dopamine neurons during reward-based learning

Summary

To understand functional role of the activity of dopamine neurons to salient stimulus in behavioral switching of action selection, we recorded the activity of dopamine neurons in an instrumental conditioning task in which monkeys made a series of behavioral decisions based on the distinct reward predictions. Brief flash of visual cue (Flash) was used as a salient stimulus to instruct the end of current block of trials and the start of new one.

We founded that dopamine neurons were responded to salient stimulus during this task. However number of responsive neurons to Flash was significantly less than that to conditioned stimulus (CS) and reinforcer. The activity to Flash was not related whether the monkeys changed behavioral selection or not at first trials of block. The activity of dopamine neurons to Flash was not changed through a few month of task training. Previously, we demonstrated that the responses of DA neurons to CS represent motivation properties. In contrast to the responses to CS, we founded that coding of saliency by dopamine neurons did not modulated by the levels of motivation, because the magnitudes of response to Flash were no correlation with reaction times to the start button.

These findings suggested that DA neurons differentially represented saliency and motivation during reward-based decision task.

Introduction

Substantial body of evidences suggested that the striatum crucially involved in reward-based learning, behavioral switching and motivational control of voluntary movements (Aosaki et al., 1994; Hollerman et al., 1998; Kawagoe et al., 1998; Shidara et al., 1998; Redgrave et al., 1999a). The midbrain dopamine (DA) neurons, located in the substantia nigra and ventral tegmental area, project to the striatum and frontal cortices. Midbrain DA system is thought to one of a central structure to provide reward and motivation related information to the striatum (Robbins and Everitt, 1996; Schultz, 1998).

Schultz and colleagues reported that DA neurons respond to unexpected reinforcer and reward associated sensory stimulus (Schultz et al., 1993; Mirenowicz and Schultz, 1994). They proposed that the activity of DA neurons to reinforcer encode prediction error of reward and provide effective reinforcement signal in reinforcement learning (Schultz et al., 1997; Hollerman and Schultz, 1998; Schultz, 1998). (Schultz et al., 1997; Hollerman and Schultz, 1998; Schultz, 1998). In accordance with this view, several studies demonstrated that responses of DA neurons to reinforcer were varied monotonically according to probability of reward in classical conditioning task (Fiorillo et al., 2003) and instrumental conditioning task (Satoh et al., 2003; Morris et al., 2004). In addition, reward-related responses of DA neurons preciously encode positive and negative reward prediction errors (Satoh et al., 2003) and show context dependency (Nakahara et al., 2004). These observations provided robust experimental supports for reinforcement learning. Furthermore, we reported that the activity of DA neurons to conditioned stimulus (CS) appears to represent motivational properties (Satoh et al., 2003).

On the other hand, several studies suggested that tonic levels of DA modulate behavioral selections and switching (Redgrave et al., 1999a). Furthermore, DA neurons also respond to biologically salient stimulus including novel, unexpected and intense sensory stimulus (Steinfels et al., 1983; Ljungberg et al., 1992; Horvitz et al., 1997; Horvitz, 2000). In addition, some populations of DA neurons increased firing rates to aversive stimulus (Chiodo et al., 1980; Schultz and Romo, 1987; Mantz et al., 1989; however Ungless et al., 2004). Based on these observations, Redgrave and colleagues hypothesized alternative functional roles of DA neurons in behavioral switching and reallocation processes (Redgrave et al., 1999b). However, it remains unclear whether the phasic activity of DA neurons to sensory stimulus related to switching of action selections or behavioral strategies.

To address this issue, we focused on the activity of DA neurons to salient stimulus during an instrumental conditioning task. In this study, we used Flash stimulus as instruction of the end of current behavioral strategy and the start of new one. To investigate whether the

responses to salient stimulus were related to behavioral switching, we analyzed relationship between the responses to Flash and behavioral switching of selection at first trials of block. To examine whether the responses of DA neurons to salient stimulus modulated by the levels of motivation, we analyzed correlation between magnitudes of Flash responses and RTs to the start button.

Materials and Methods

General

Same Japanese monkeys (*Macacca fuscata*: monkey DN and SK) were used as in previous study (Satoh et al., 2003). 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. Details of surgical, data acquisition procedures were described previously (Satoh et al., 2003).

Behavioral task

The apparatus and behavioral task were same as in previous study. The monkeys were trained to sit on a primate chair with head restrained in a dimly lighted and sound attenuated room. In front of the animals, a panel equipped with rectangular button with red LED (start LED, 14x14 mm) at the bottom, three push buttons with green LED (target LED, 14x14 mm) in the middle row and red LED (GO LED, 4 mm diameter) was placed. For facilitating behavioral training, the monkeys were initially trained to behavioral association of a high-tone beep (1kHz for 100ms) with reward water. A task trial started with illumination of the start LED on the push button (Fig. 1A). After the monkeys continued to depress the start button for 400 ms, start LED was turned off and the target LEDs and GO LED were illuminated. If the monkeys maintained to hold the start button for variable delay periods (600, 700 or 800 ms), the GO LED was turned off. Then the monkeys were required to release the start button within 1s and depress the one of three target buttons. If the incorrect button was depressed, a low-tone beep (300Hz for 100 ms) occurred with a delay of 500ms. The Next trial began with illumination of start LED at 7.5 s after releasing push button. If the correct button was depressed, a high-tone beep (1kHz for 100 ms) occurred after 500ms. Small amount of reward water was delivered through a spout in front of the monkey's mouth. If the monkeys released the start button before the end of delay period, we counted this type of error as an early release error. If the monkeys did not responded within 1s after the offset of GO LED, we counted this as a late error.

One of three target buttons was used as the correct button through a single block of trials. Because location of correct button changed unpredictably in each block, the monkeys searched for the correct button on a trial and error manner. If the monkeys found the correct button, they received a reward three times by selecting the same button during three constitutive trials. Thus trials in a single block were divided into two epochs (Fig. 1B). The first epoch was the trial-and-error epoch. Three types of trials occurred: trials in which the

monkeys selected the correct button at first, second and third trial in a single block (N1, N2 and N3, respectively). The second was the repetition epoch. Two types of trials occurred: the first and second trials in the repetition epoch (R1 and R2, respectively). The amount of reward was 0.35 ml in the trial and error epoch, and 0.25 ml in repetition epoch.

For instruction of the end of current block of trials and the transition to new block, three target LEDs were simultaneously flashed (Flash) for 100 ms at 2 s after releasing push button (Fig. 1C). New block of trials began with illumination of the start LED after 3.5 s.

We divided the experimental sessions (7 months in monkey DN, 3 months in monkey SK) into an early stage, partially learned stage, and later, fully learned stage, based on the difference in task performance (Satoh et al., 2003).

Data acquisition and analysis

Details of data acquisition procedures were described previously (Satoh et al., 2003). The activity of single neurons was recorded extracellularly with epoxy-coated tungsten microelectrodes (exposed tips of 15μm length and impedances of 2-5MΩ; 26-10-2L, Frederic Hear, Bowdoinham). The electrodes were inserted through a stainless steel guide tube (OD 650μm) and advanced with an oil-drive micromanipulator (MO-95, Narishige, Tokyo). Single unit was discriminated by a spike sorter with a template-matching algorism (MSD4, Alpha Omega, Nazare). The times of the action potentials and the onset and offset of behavioral events were recorded on a laboratory computer with time resolution of 1 ms. Responsiveness of neurons to a task event was evaluated by comparing the discharge rate during the 50 ms (5 bins) time window with that during 250 ms (25 bins) control window just before the onset of it (Kimura, 1986). The time window was moved in step from the onset of an event. Only neurons with significant changes, at least three consecutive comparisons ($p < 0.05$, the Wilcoxon signed rank test), were counted as responsive. The onset and offset of response were determined as the beginning and end of significant changes of activity. To quantify the neural response, we selected a fixed time period for phasic responses determined by the average onset and offset response latencies and measured the firing frequencies from baseline discharge in each neuron. The baseline discharge rates were calculated as the average discharge rate during the 1 sec before the onset of start LED.

To investigate relation to behavioral switching of action selection and neuronal responses, we calculated the average change rates of action selection in N1 trials. Location of the correct button changed unpredictably in each block. If the monkeys selected the button in N1 trials that bad been uncorrected in previous block of trials, we counted that as a changed selection trial.

Results

We recorded the activity of DA neurons characterized by histological and/or physiological properties in two monkeys. We reanalyzed same populations of neuronal data (52 DA neurons from monkey DN; 56 DA neurons from monkey SK) described in previous study (Satoh et al., 2003).

Responses to Flash stimulus

Previously, we reported that about half of DA neurons significantly responded to CS (27/52 neurons in monkey DN; 27/56 neurons in monkey SK) and reinforcers (32/52 neurons in monkey DN; 38/56 in monkey SK). Although the rest of task events such as GO LED, hand movement and reward did not elicit significant modulation of DA neuron activity, some populations of DA neurons responded significantly to the Flash stimulus presented at the transition period between two blocks of trials (Table1, 1/52 neurons in monkey DN; 14/56 neurons in monkey SK, $P < 0.05$, Wilcoxon test). These results are based on the neuronal activity both in the partially learned and fully learned stages. Fig 1A shows example response of DA neurons of monkey SK to Flash. A short burst of discharge was occurred after the instruction of Flash. Population histogram of 56 DA neurons of monkey SK in Fig. 1C demonstrated an increase of discharge rate to Flash, but that of monkey DN in Fig. 1B did not show modulation of activity. Average magnitude of responses to the Flash was significantly increased from baseline discharge in monkey SK (1.46 ± 0.2, mean ± SE; $p < 0.01$) (Fig. 1D).

What is the functional role of responses of DA neurons to Flash? Flash was presented as a salient stimulus to instruct the end of current block and start new one. In contrast to CS, Flash itself was not associated with reward and any reactions to it were not required. To investigate whether the activity of DA neurons to saliency stimulus was related to behavioral switching of action selections, we calculated the average change rates of action selection in N1 trials. Two monkeys behaved differentially. Monkey DN changed their selections in half of N1 trials (average change rate: 48.7 ± 20.2 %, mean ± SD), but monkey SK tended to chooses same button that had been a correct one in previous block of trials (average change rate: 1.5 ± 2.3 %, mean ± SD) (Fig. 3). In addition, behavioral switching of action selection at N1 trials and responses to CS, was not modulated the activity of DA neurons to CS ($p > 0.3$ for monkey DN, Wilcoxon rank test).

The response to the Flash was present throughout two learning stages in monkey SK. Number of responsive neurons of monkey SK to Flash was not significantly changed during development of learning (Table1, 7/19 in partially learned stage; 7/37 in fully learned stage; $p > 0.1$, $\chi 2 = 2.15$). Fig. 4A shows the ensemble averages of activity of monkey SK to Flash in

partially learned and fully leaned stages. DA neurons produced a brisk response to Flash in both learning stages. Magnitude of Flash responses was insignificant difference between two learning stages (Fig. 4B, $p > 0.2$, Mann-Whitney U test).

Differential coding of saliency and motivation

Previously, we demonstrated that the responses of DA neurons to CS were negatively correlated with the reaction times (RTs) to the start button. Because RTs were used as behavioral measures of the level of motivation, responses of DA neurons to CS represented motivational properties. To elucidate relationship between the levels of motivation and RTs to the start button in this task, we divided performance related errors into two types. The early release error occurred when the monkey released the start button before the end of delay period. The late error occurred when the monkey did not responded within 1s after the offset of GO LED. Both types of errors were mainly occurred in trial and error epoch (Fig. 5). Rates of early release error were significantly changed by RTs to the start button in trial and error epoch (Fig. 5A; $p < 0.01$, $\chi 2 = 12.0$ in monkey DN; $p < 0.001$, $\chi 2 = 22.8$ in monkey SK). In monkey SK, Rates of late error were significantly changed by RTs to the start button in trial and error epoch (Fig. 5B; $p < 0.05$, $\chi 2 = 8.8$ in monkey SK). These observations supported a view that RTs to the start button may reflect the level of motivation to work for a reward.

CS and Flash were same modality (visual) of sensory stimulus in this study. Basic response properties, such as response latencies and durations, were not significantly different between CS and Flash in monkey SK (average onset latencies: $p > 0.09$, unpaired t-test; average response durations: $p > 0.9$, unpaired t-test). However DA neurons responded to CS and Flash differentially. Number of responsive DA neurons to Flash was significantly less than that to CS ($p < 0.001$, $\chi 2 = 33.0$ in monkey DN; $p< 0.05$, $\chi 2 = 6.5$ in monkey SK) and to reinforcer ($p < 0.001$, $\chi 2 = 42.6$ in monkey DN; $p < 0.001$, $\chi 2 = 13.1$ in monkey SK) in both monkeys. Fig. 6 demonstrated that the activity of DA neurons to Flash was no correlation with RTs to the start button in monkey SK ($r = -0.04$, $p > 0.5$). This observation was in sharp contrast to previous result in which the response to CS was significantly negative correlation with RTs to the start button (Sato et. al., 2003).

Most of Flash responsive neurons were also significantly responded to CS (Table1, 1/1 neurons in monkey DN; 11/14 neurons in monkey SK). To examine a possibility that only neurons responded to both CS and Flash showed similar response properties to CS and Flash, we analyzed a correlation between RTs to the start button and responses to the Flash from 11 neurons responded to both CS and Flash in monkey SK. However there was no correlation between them ($r = -0.05$, $p > 0.7$).

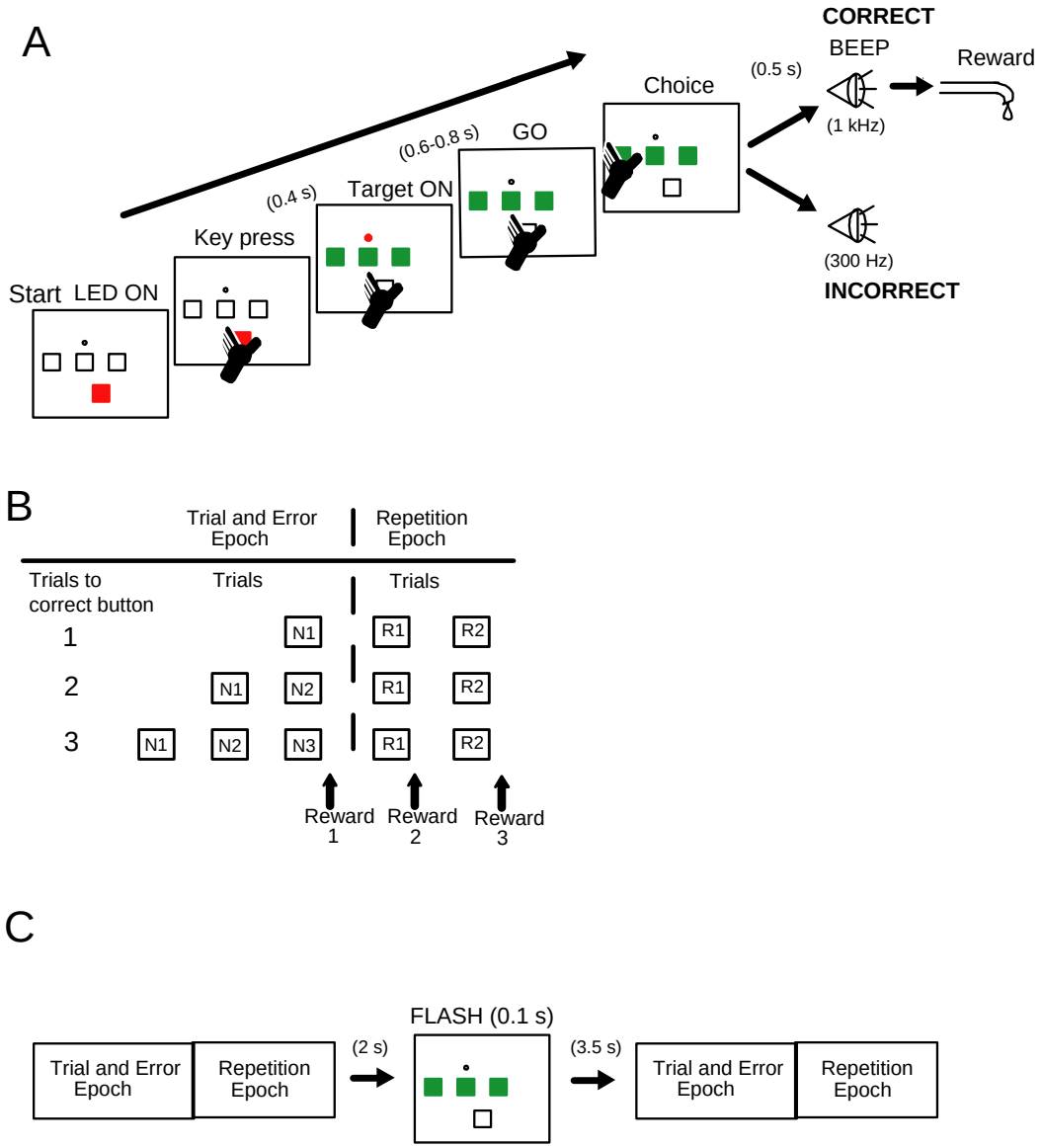

Figure 1. Schematic illustrations of behavioral task *A*, Temporal sequences of task events in a single trial. See details in Materials and Methods. *B*, Trials in a single block is consisted with two epochs (Trial and Error Epoch and Repetition Epoch) and 5 trial types (N1, N2, N3, R1, R2) defined by the bases of correct and incorrect button choices. *C*, Illustration of temporal sequence of Flash stimulus presented at the transition period between two blocks of trials.

Table 1. Number of Responsive DA Neurons to Flash, Conditioned Stimulus (CS) and Reinforcers

Monkey DN		
	Partially Learned Stage	Fully Learned Stage
Flash	1	0
Conditioned Stimulus (CS)	4	23
Reinforcers	2	30
CS and Flash	1	0
Reinforcers and Flash	1	0
CS and Reinfoecer	1	18
CS, Reinforcers and Flash	1	0
Total	8	44

Monkey SK		
	Partially Learned Stage	Fully Learned Stage
Flash	7	7
Conditioned Stimulus (CS)	11	16
Reinforcers	12	26
CS and Flash	7	4
Reinforcers and Flash	6	5
CS and Reinforcer	7	16
CS, Reinforcers and Flash	6	3
Total	19	37

Figures are number of responsive neurons determined by Wilcoxon single rank test at $p<0.05$ (Kimura, 1986).

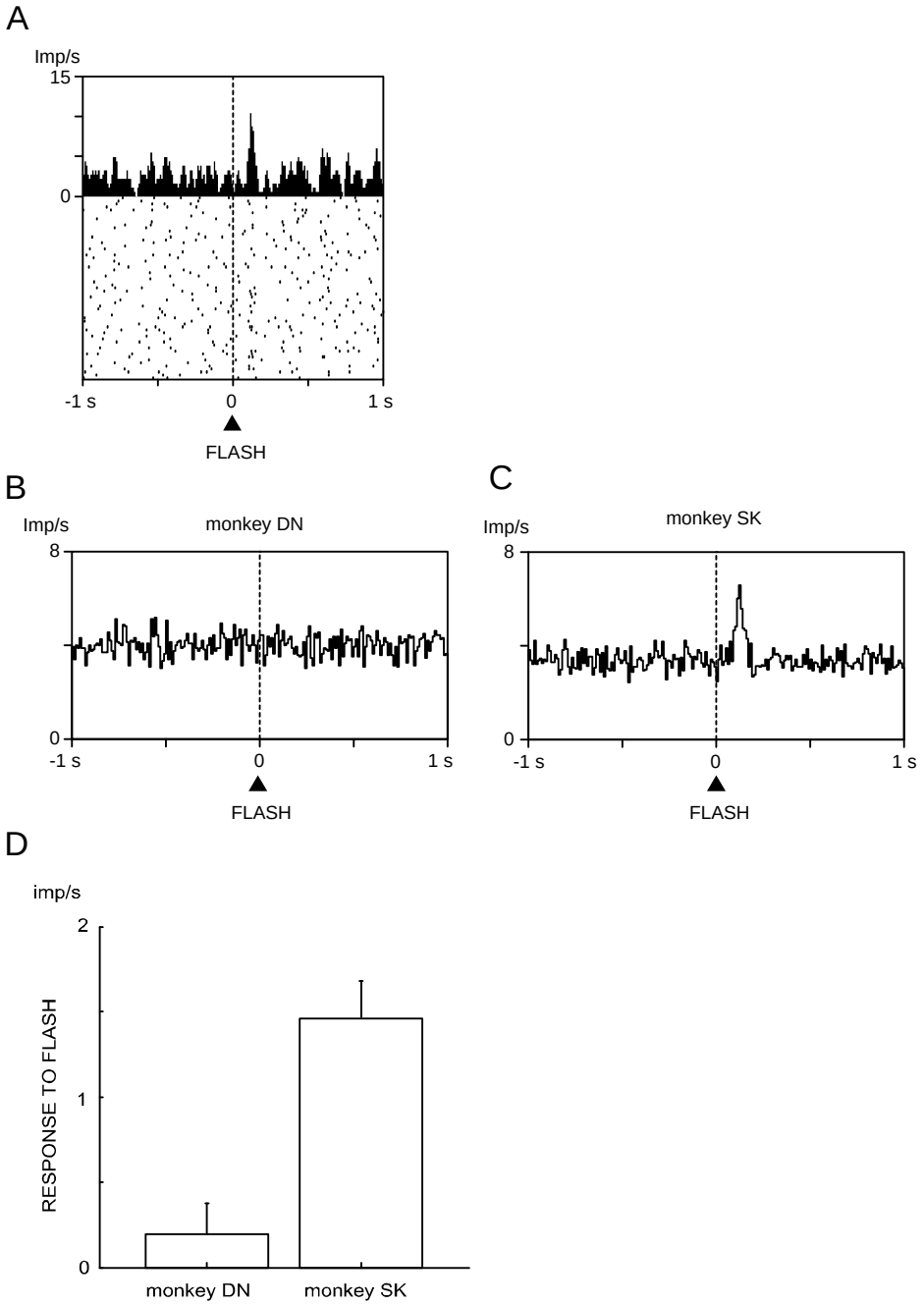

Figure 2. Response of DA neurons to the Flash stimulus presented at the end of correct R2 trials *A*, A single DA neurons of activity of monkey SK to Flash. In the raster plot, each dot represents the time of an impulse and each row indicates a trial. Raster display and peri-event histogram are aligned on the onset of Flash (vertical interrupted line). Histogram is constructed by summation of impulses and displayed as impulses per second. Bin width = 10 ms. *B*, Population histogram of 50 DA neurons of monkey DN to Flash. Bin width = 10 ms. *C*, Same as (B) but for 56 DA neurons of monkey SK. *D*, Average magnitude of response to Flash (Mean ± SEM) during fixed time window 80-180 ms after the onset of Flash relative to

the baseline discharge rate in 2 monkeys.

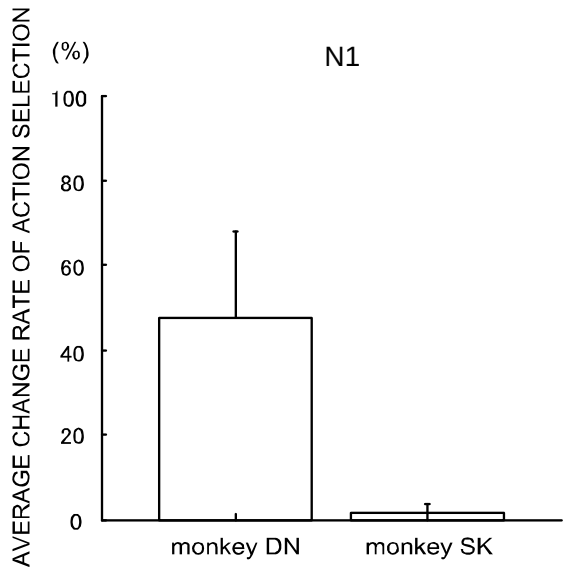

Figure 3. Average change rate (Mean ± SD) of action selection at N1 trials in two monkeys.

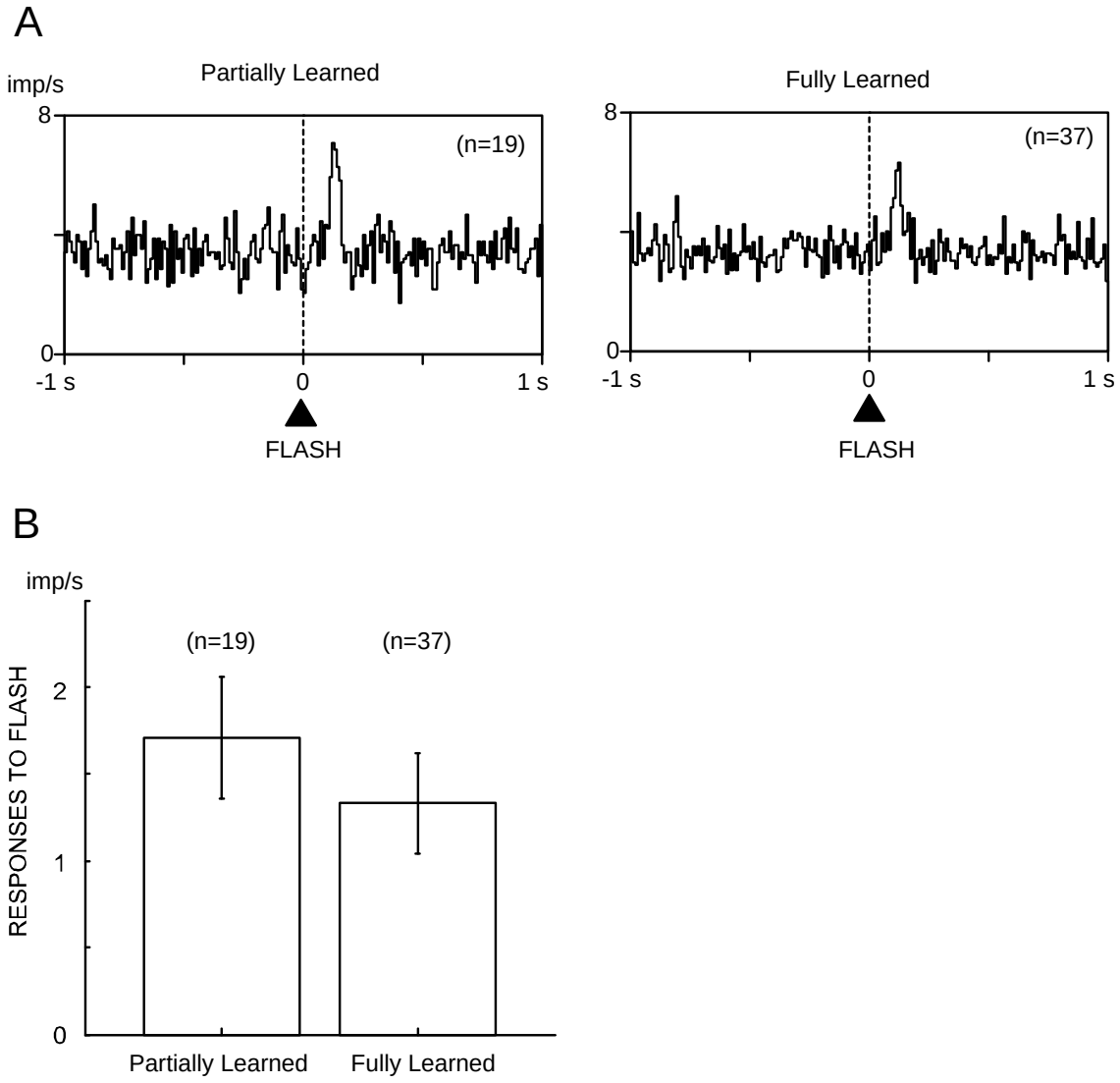

Figure 4. Responses of monkey SK to Flash in the partially learned and fully learned stage. *A*, Comparisons of population response histograms to the Flash between partially learned (*left*) and fully learned stage (*right*). Population histograms are aligned on the onset of Flash (vertical interrupted line). Number of neurons included for each histogram is shown in parentheses. Bin width is 10 ms. *B*, Average magnitude of response of monkey SK to Flash during fixed time window (80-180 ms) relative to the baseline discharge rate. Bars indicate standard errors.

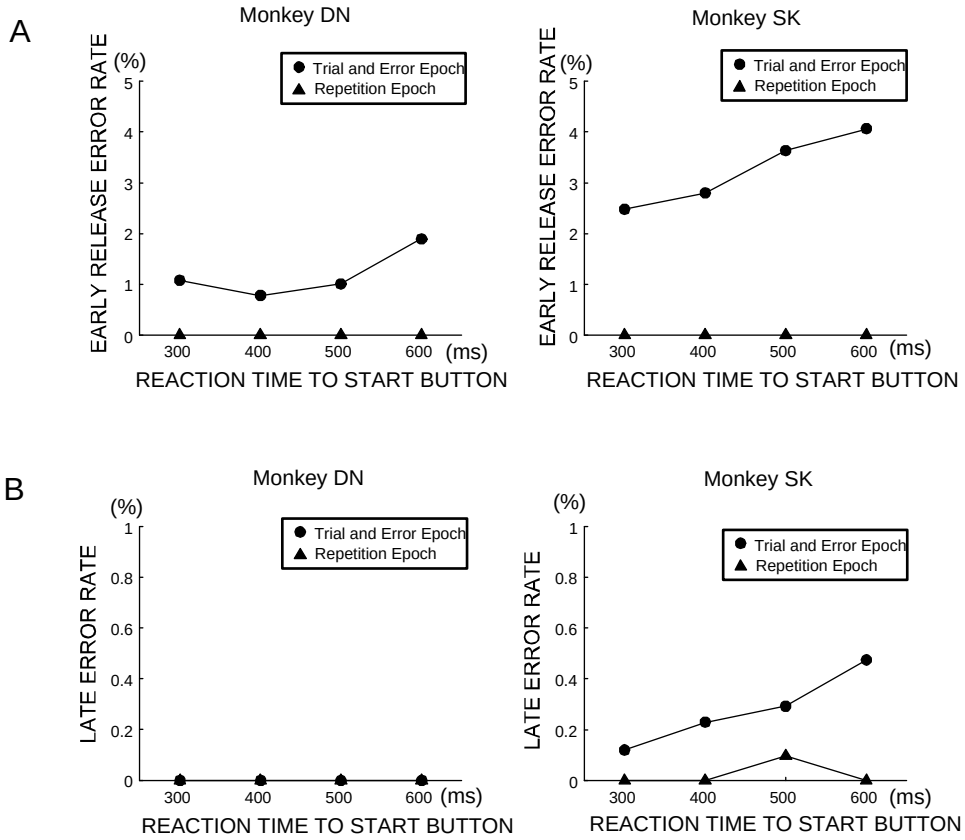

Figure 5. Analysis of task performance against reaction times to the start button in two monkeys *A*, Plots of percentage of early release errors as a function of RTs to depress the start button after CS in trial and error (filled circle) and repetition (filled triangle) epoch. *B*, Plots of percentage of late errors as a function of RTs to depress the start button after CS in trial and error (filled circle) and repetition (filled triangle) epoch.

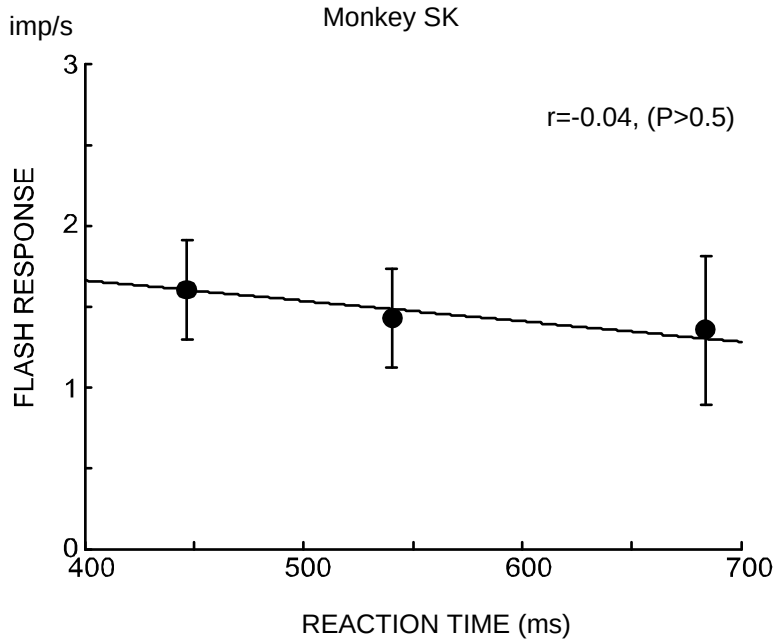

Figure 6. Scatter plot of the average response to Flash (mean ± SEM) and reaction times to the start button in monkey SK. RTs were divided into 3 groups based on RTs. Regression line is superimposed.

Discussion

Relation to behavioral switching of action selection

Previous studies reported that DA neurons respond to biologically salient stimulus including novel, unexpected and intense sensory stimulus (Steinfels et al., 1983; Ljungberg et al., 1992; Horvitz et al., 1997; Horvitz, 2000). It has been hypothesized that activities of DA neurons to biologically salient stimulus were related to behavioral switching and/or reallocation processes (Redgrave et al., 1999b). However, there were no studies to address whether the activity of DA neurons to saliency was related to behavioral switching of selections. In this study, Flash was used as a salient stimulus to instruct the end of current block of trials and the start of new one. In sharp contrast to CS, Flash itself was not associated with reward and any reactions were not required. In both monkeys, Reaction times to the start button at the first trials of block (N1 trials) were significantly longer than those at the final trials of block (R2 trials) in fully learned stages (Satoh et al., 2003). Behavioral differences between N1 and R2 trials were developed through the initial and late stages of learning. These observations suggested that the monkeys could be recognized the change of block of trials.

In this study, we founded that DA neurons were responded to Flash stimulus (1/52 neurons in Monkey DN; 14/56 neurons in monkey SK) during reward-based decision task. Redgrave and colleagues proposed the behavioral switching hypothesis of DA function that increase of DA activity facilitates behavioral switching and decrease of DA activity suppress behavioral switching (Redgrave et al., 1999b). However, the response properties of DA neurons to the Flash did not consisted with their behavioral switching hypothesis. Although Monkey DN changed selection of target button at about half of N1 trials, Flash did not produced increase of discharge rate. Although Flash produced a robust response in monkey SK, Monkey SK did not change their selection at most of N1 trials. In addition, the response to the Flash was maintained throughout two learning stages in monkey SK. The ability of DA neurons to response to behavioral salient stimulus, such as unexpected, novelty and intensity stimulus, may be important to trigger behavioral switching of selections. However the response to salient stimulus were not related whether the monkeys changed behavioral selection or not.

Current block of trials was ended when the monkeys received three rewards in each block. Because nonhuman primates have numerical representation (Sawamura et al., 2002; Nieder and Miller, 2004), there was a possibility that monkeys could be recognized the change of block independent of the instruction of Flash. If this possibility was true, we should not count Flash stimulus as the instruction to terminate current block of trials and start new one. Therefore, to confirm behavioral switching hypothesis of DA function, future studies were

required to introduce a task that a salient stimulus was a key factor to determine behavioral switching of selection or strategy.

Differential coding of saliency and motivation

DA neurons responded to multiple task events such as CS, reinforcer and salient stimulus in this task. Previously, we demonstrated that the responses of DA neurons to CS represent motivational properties to work for reward, because magnitudes of responses to CS were negatively correlated with RTs to the start button. Furthermore, the magnitude of response to CS was positively correlated with that to reinforcer. These results suggested that the levels of motivation modulated the responses of DA neurons to reinforcer. On the other hands, the magnitudes of responses to Flash were no correlation with RTs to the start button in monkey SK. This suggested that the levels of motivation did not modulate the coding of saliency by DA neurons. These results suggested that DA neurons coded saliency and motivation differentially during reward-based decision task.

Reference

Aosaki T, Graybiel AM, Kimura M (1994) Effect of the nigrostriatal dopamine system on acquired neural responses in the striatum of behaving monkeys. Science 265:412-415.

Chiodo LA, Antelman SM, Caggiula AR, Lineberry CG (1980) Sensory stimuli alter the discharge rate of dopamine (DA) neurons: evidence for two functional types of DA cells in the substantia nigra. Brain Res 189:544-549.

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

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

Hollerman JR, Tremblay L, Schultz W (1998) Influence of reward expectation on behavior-related neuronal activity in primate striatum. J Neurophysiol 80:947-963.

Horvitz JC (2000) Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. Neuroscience 96:651-656.

Horvitz JC, Stewart T, Jacobs BL (1997) Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. Brain Res 759:251-258.

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

Kimura M (1986) The role of primate putamen neurons in the association of sensory stimuli with movement. Neurosci Res 3:436-443.

Kobayashi Y, Inoue Y, Yamamoto M, Isa T, Aizawa H (2002) Contribution of pedunculopontine tegmental nucleus neurons to performance of visually guided saccade tasks in monkeys. J Neurophysiol 88:715-731.

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

Mantz J, Thierry AM, Glowinski J (1989) Effect of noxious tail pinch on the discharge rate of mesocortical and mesolimbic dopamine neurons: selective activation of the mesocortical system. Brain Res 476:377-381.

Mirenowicz J, Schultz W (1994) Importance of unpredictability for reward responses in primate dopamine neurons. J Neurophysiol 72:1024-1027.

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.

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

Nieder A, Miller EK (2004) A parieto-frontal network for visual numerical information in the monkey. Proc Natl Acad Sci U S A 101:7457-7462.

Redgrave P, Prescott TJ, Gurney K (1999a) The basal ganglia: a vertebrate solution to the selection problem? Neuroscience 89:1009-1023.

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

Robbins TW, Everitt BJ (1996) Neurobehavioural mechanisms of reward and motivation. Curr Opin Neurobiol 6:228-236.

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.

Sawamura H, Shima K, Tanji J (2002) Numerical representation for action in the parietal cortex of the monkey. Nature 415:918-922.

Schultz W (1998) Predictive reward signal of dopamine neurons. J Neurophysiol 80:1-27.

Schultz W, Romo R (1987) Responses of nigrostriatal dopamine neurons to high-intensity somatosensory stimulation in the anesthetized monkey. J Neurophysiol 57:201-217.

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

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

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.

Steinfels GF, Heym J, Strecker RE, Jacobs BL (1983) Behavioral correlates of dopaminergic unit activity in freely moving cats. Brain Res 258:217-228.

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

Ungless MA, Magill PJ, Bolam JP (2004) Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. Science 303:2040-2042.

Summary and General Discussion

In this thesis, I recorded the activity of DA neurons during an instrumental conditioning task that monkeys made a series of behavioral decision based on trial-specific reward expectations. I found that about 50% (27/52 in monkey DN, 27/56 in monkey SK) and 60% (32/52 in monkey DN, 38/56 in monkey SK) of DA neurons were significantly responsive to CS and reinforcer. In addition, small population of DA neurons (1/52 in monkey DN, 14/56 in monkey SK) responded to salient stimulus presented as instruction to terminating current block of trials. In chapter 1 of this thesis, I focused on the activity of DA neurons to CS and reinforcers. I found that the magnitude of responses to CS represents motivational properties. In addition, I revealed that neuronal responses to reinforcer quantitatively encode REEs during an instrumental conditioning task. Coding of REEs was not modulated by which of three target buttons monkeys selected. These finding suggested that REEs signals of DA neurons to reinforcer act as general teaching signals during reward-based learning. In addition, about a half of DA neurons (19/52 in monkey DN; 23/56 in monkey SK) were significantly responsive both CS and reinforcer. I demonstrated that magnitudes of neuronal responses to CS were positively correlated with those to reinforcer, suggesting modulation of ability of general teaching signals by motivational level. In addition, the gain of coding REEs developed through task learning, while coding of motivational properties remained consistent during the learning. In chapter 2 of this thesis, I focused on the activity of DA neurons to salient stimulus. I founded that small populations of DA neurons responded to salient stimulus. The responses of DA neurons to Flash were not related to behavioral switching of selection. In addition, the responses to Flash were maintained through the initial and late stages of learning. I demonstrated that magnitudes of responses to Flash were no correlation with RTs to the start button, suggesting differential coding of saliency and motivation by DA neurons.

Coding of reward prediction errors as general teaching signals

Reinforcement learning theories are computer frameworks that an agent tries to maximize cumulative sum of rewards through adaptation of its behaviors. These theories proposed that an agent adapt its behavior based on reward prediction errors that act as general teaching signals (Barto, 1995). A large number of reinforcement learning models proposed that the basal ganglia circuits are best candidates for implementation of these computer algorisms (Barto, 1995; Montague et al., 1996; Suri and Schultz, 1998). The basal ganglia received topographically organized inputs from wide cortical areas and attention-related signals from CM/Pf complex of thalamus (Minamimoto and Kimura, 2002). Previous studies

hypothesized that modulation inputs from DA neurons act as general reinforcement signals. In chapter 1 of this thesis, I demonstrated that the responses of DA neurons to reinforcer precisely represent REEs. This finding was consistent with series of studies by Schultz and colleagues (Mirenowicz and Schultz, 1994; Hollerman and Schultz, 1998; Fiorillo et al., 2003). But, we revealed, for the first time, that responses of DA neurons to reinforcer encode positive and negative REEs quantitatively in a series of reward expectation-based decision process in the instrumental conditioning task. In addition, there were no significant selectivity between neuronal responses to reinforcer separated based upon monkey decisions. These findings suggested that quantitative representation of REEs signals by DA neurons were suitable for general reinforcement signals in reward-based learning. REEs signals of DA neurons must effective to adaptive changes of synaptic transmission in striatum and prefrontal cortices through dopamine dependent synaptic plasticity (Wickens et al., 1996; Otani et al., 1999; Calabresi et al., 2000; Reynolds et al., 2001). Learning related activity changes of striate neurons (Aosaki et al., 1994; Tremblay et al., 1998) also support this view.

Dual coding of motivation and reinforcement properties

Although most of reinforcement learning theories less mentioned about motivation, several computational models of learning, recently, suggested importance of motivation (Dayan and Balleine, 2002; McClure et al., 2003) and attention (Dayan et al., 2000) in reward related learning and proposed revised algorithms of models. Psychological theories suggested that motivation has several important effects on reward related learning and decision-making. First, Pavlovian CS has evocative effect to conditioned responses to reward (Lovibond, 1983; Gawin, 1991) and acts as conditioned reinforcer. Motivational states modulate these ability of pavlovian incentive cue (Cardinal et al., 2002). Second, motivational manipulations such as satiation and deprivation modulate the tasty of reward (Berridge, 2001) and the effectiveness as reinforcer (Skinner, 1953; Michael, 1982, 2000). Pharmacological studies suggested that DA system plays a pivotal role for Pavlovian CS and reinforcer to cause behavioral changes (Wise et al., 1978; Dickinson et al., 2000; Wyvell and Berridge, 2000; Di Ciano et al., 2001). However, there was no experimental neural evidence to link motivation and reinforcement mechanisms.

In chapter 1 of this thesis, I found that about 50% (27/52 in monkey DN, 27/56 in monkey SK) of DA neurons were significantly responsive to CS. Although the responses to CS were varied with trial types, these were not accurately estimated by the expectation of reward as a reward probability. However, I demonstrated, for the first time, that the magnitudes of responses to CS have significant negative correlation with behavioral reaction times to CS. Because reaction times are one of behavioral measures reflecting levels of

motivation (Shidara et al., 1998; Kobayashi et al., 2002), this finding suggested that responses of DA neurons to CS represent motivational properties. Interestingly, Injection of DA agonist enhanced the evocative effect of Pavlovian CS (Wyvell and Berridge, 2000). On the other hand, administration of DA antagonist reduced the effectiveness of Pavlovian CS (Dickinson et al., 2000), CS-elicited reward-seeking behavior (Di Ciano et al., 2001; Yun et al., 2004) and sexually conditioned incentives (Lopez and Ettenberg, 2002). These studies are consistent with a view that motivational properties of DA neurons activity to CS are one possible candidate to modulate the ability of Pavlovian CS.

I found that about a half of DA neurons (19/52 in monkey DN; 23/56 in monkey SK) were significantly responsive both CS and reinforcer. It is clear that learning rates and task performances are influenced by motivational manipulation (Skinner, 1953). To elucidate the functional integration of incentive motivation and reinforcement coding on single DA neuron, I investigated relationship between motivational properties attributed to CS and REEs to reinforcer. It was found that magnitudes of responses to CS were positively correlated with those to reinforcer at trials with the same levels of reward expectation, same trial type. This result suggested that the effectiveness of REEs as general teaching signal is modulated by motivational level. Recently, several computational models of reinforcement theories pointed out the importance of motivation in reward related learning (Dayan and Balleine, 2002; McClure et al., 2003). This result suggested new model of reinforcement learning than currently proposed.

Furthermore, I examined the development of responses to CS reflecting motivational properties and gain of coding REEs, number of DA neuron spikes encoding REEs, through learning process of task rules. I found that the gain of motivation coding did not change through initial and late learning stage, but the gain of coding REEs developed through task learning. Before the start of training of instrumental conditioning task, I used forward chaining procedure to help monkeys acquire several steps of behavioral chaining. Initially, First step of stimulus presented and a correct response was established by reward. After the response of first step was learned, second step of action was reinforced by reward. This training procedure was important to effectively acquire several steps of behavioral chaining, especially reward was delivered only at final step (Malott and Suarez, 2003). Pavlovian incentive theories suggested that Pavlovian CS, which is associated with reinforcer, has incentive effect on goal-directed and reward seeking behaviors (Cardinal et al., 2002). Through forward chaining procedure, CS presented in the first step was associated with reward and may have evocative effects to conditioned response. This process must play a fundamental but distinct role in acquiring behavioral chaining effectively through reinforcement process.

Differential coding of saliency and motivation

In this thesis, I founded that DA neurons responded to several task events such as CS, outcomes of reinforcer and salient stimulus during reward-based learning. Although half of DA neurons coded both motivation and REEs, small population of DA neurons represented saliency during reward-based learning. In addition, the levels of motivation did not modulate the responses of DA neurons to salient stimulus, suggesting differential coding of saliency and motivation by DA neurons. Previous studies suggested that DA neurons received information of saliency from superior colliculus or raphe nucleus (Comoli et al., 2003; Schultz, 1998). The ability of DA neurons to response to behavioral salient stimulus, such as unexpected, novelty and intensity stimulus, may be important to trigger behavioral switching of selections. However the response to salient stimulus were not related whether the monkeys changed behavioral selection or not. DA neurons provided saliency independent of the levels of motivation.

References for Summary and General Discussion

Aosaki T, Graybiel AM, Kimura M (1994) Effect of the nigrostriatal dopamine system on acquired neural responses in the striatum of behaving monkeys. Science 265:412-415.

Barto AG (1995) Adaptive critics and the basal ganglia. In: Models of Information Processing in the Basal Ganglia (Houk JC, Davis JL, Beiser DG, eds), pp 215-232. Cambridge: The MIT Press.

Berridge KC (2001) REWARD LEARNING: Reinforcement, Incentives, and Expectations. In: The Psychology of Learning and Motivation, Volume 40, pp 223-278. San Diego: Academic Press.

Calabresi P, Gubellini P, Centonze D, Picconi B, Bernardi G, Chergui K, Svenningsson P, Fienberg AA, Greengard P (2000) Dopamine and cAMP-regulated phosphoprotein 32 kDa controls both striatal long-term depression and long-term potentiation, opposing forms of synaptic plasticity. J Neurosci 20:8443-8451.

Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci Biobehav Rev 26:321-352.

Comoli E, Coizet V, Boyes J, Bolam JP, Canteras NS, Quirk RH, Overton PG, Redgrave P (2003) A direct projection from superior colliculus to substantia nigra for detecting salient visual events. Nat Neurosci 6:974-980.

Dayan P, Balleine BW (2002) Reward, motivation, and reinforcement learning. Neuron 36:285-298.

Dayan P, Kakade S, Montague PR (2000) Learning and selective attention. Nat Neurosci 3 Suppl:1218-1223.

Di Ciano P, Cardinal RN, Cowell RA, Little SJ, Everitt BJ (2001) Differential involvement of NMDA, AMPA/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of pavlovian approach behavior. J Neurosci 21:9471-9477.

Dickinson A, Smith J, Mirenowicz J (2000) Dissociation of Pavlovian and instrumental incentive learning under dopamine antagonists. Behav Neurosci 114:468-483.

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

Gawin FH (1991) Cocaine addiction: psychology and neurophysiology. Science 251:1580-1586.

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

Kobayashi Y, Inoue Y, Yamamoto M, Isa T, Aizawa H (2002) Contribution of

pedunculopontine tegmental nucleus neurons to performance of visually guided saccade tasks in monkeys. J Neurophysiol 88:715-731.

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

Lopez HH, Ettenberg A (2002) Sexually conditioned incentives: attenuation of motivational impact during dopamine receptor antagonism. Pharmacol Biochem Behav 72:65-72.

Lovibond PF (1983) Facilitation of instrumental behavior by a Pavlovian appetitive conditioned stimulus. J Exp Psychol Anim Behav Process 9:225-247.

Malott RW, Suarez EAT (2003) Principles of behavior. New York: Person Prentice Hall.

McClure SM, Daw ND, Montague PR (2003) A computational substrate for incentive salience. Trends Neurosci 26:423-428.

Michael J (1982) Distinguishing between discriminative and motivational functions of stimuli. J Exp Anal Behav 37:149-155.

Michael J (2000) Implications and refinements of the establishing operation concept. J Appl Behav Anal 33:401-410.

Minamimoto T, Kimura M (2002) Participation of the thalamic CM-Pf complex in attentional orienting. J Neurophysiol 87:3090-3101.

Mirenowicz J, Schultz W (1994) Importance of unpredictability for reward responses in primate dopamine neurons. J Neurophysiol 72:1024-1027.

Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16:1936-1947.

Otani S, Auclair N, Desce JM, Roisin MP, Crepel F (1999) Dopamine receptors and groups I and II mGluRs cooperate for long-term depression induction in rat prefrontal cortex through converging postsynaptic activation of MAP kinases. J Neurosci 19:9788-9802.

Reynolds JN, Hyland BI, Wickens JR (2001) A cellular mechanism of reward-related learning. Nature 413:67-70.

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

Schultz W (1998) Predictive reward signal of dopamine neurons. J Neurophysiol 80:1-27.

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.

Skinner BF (1953) Science and human behavior. New York: Macmillan.

Suri RE, Schultz W (1998) Learning of sequential movements by neural network model with dopamine-like reinforcement signal. Exp Brain Res 121:350-354.

Tremblay L, Hollerman JR, Schultz W (1998) Modifications of reward expectation-related neuronal activity during learning in primate striatum. J Neurophysiol 80:964-977.

Wickens JR, Begg AJ, Arbuthnott GW (1996) Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. Neuroscience 70:1-5.

Wise RA, Spindler J, deWit H, Gerberg GJ (1978) Neuroleptic-induced "anhedonia" in rats: pimozide blocks reward quality of food. Science 201:262-264.

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

Yun IA, Wakabayashi KT, Fields HL, Nicola SM (2004) The ventral tegmental area is required for the behavioral and nucleus accumbens neuronal firing responses to incentive cues. J Neurosci 24:2923-2933.

Bibliography

Research Papers

Satoh T., Nakai S., Sato T. and Kimura M. (2003) Correlated coding of motivation and outcome of decision by dopamine neurons. *Journal of Neuroscience*, 23: 9913-9923.

Kimura M., Matsumoto N., Okahashi K., Ueda Y., Satoh T., Minamimoto T., Sakamoto M. and Yamada H. (2003) Goal-directed,serial and synchronous activation of neurons in the primate striatum. *NeuroReport*, Vol14 :799-802.

Kimura M., Matsumoto N., Ueda Y., Satoh T., Minamimoto T., Yamada H. (2002) Involvement of the basal ganglia and dopamine system in learning and execution of goal-directed behavior. *Adv Behav Biol*, 377-380.

Abstracts

Satoh T., Nakai S., Kimura M. (2002) Dopamine neurons encode teaching signals for trial-and-error behavioral decision learning. *Neurosci. Res. Suppl*, 26: S88.

Satoh T., Kimura M. (2001) Activity of primate midbrain dopamine neurons during trial and error problem solving tasks. *Neurosci. Res. Suppl*, 25: S201.

Satoh T., Kimura M. (1998) Activity of midbrain dopamine neurons during learning sequential push button tasks in monkeys. *Soc. Neurosci. Abst*. 24: 1651.

Satoh T., Kimura M. (1998) Activity of midbrain dopamine neurons during learning sequential motor tasks in monkeys. *Neurosci. Res. Suppl*. 22: S235.

Acknowledgements

I would like to express my deep gratitude to Professor Minoru Kimura of Kyoto Prefectural University of Medicine for continuous encouragement, instruction and invaluable discussions from beginning of my research carrier. I would like to express my gratitude to Youichi Kumota, Sadamu Nakai and Tatsuo Sato for actively participating in this work. I wish to express my grateful to Dr. Naoyuki Mastumoto, Dr. Katushige Watanabe, Dr. Kazuyuki Samejima, Dr. Yasumasa Ueda and Dr. Takafumi Minaimoto for support and instruct me many things. I also wish to express my grateful to Dr. Yoshio Imahori for his assistant in surgery. I wish to express my thanks to the following people for support and assistance me and my work: Dr. Shuji Higashi, Dr. Hitoshi Inokawa, Hiroshi Yamada, Yukiko Hori, Hiromi Takemoto, Kenji Okahashi, Masashi Sakamoto, Sachiyo Fujioka, Ryoko Sakane and Harue Mastuda. I also wish to express my grateful to Professor Yoshio Hata, Dr. Satosi Ichisaka and Dr. Hiroshi Maki, who are staffs of Faculty of Medicine, Tottori University for support and encouragement. I am grateful to Professor Fujio Murakami, Professor Taishin Nomura and Professor Nobuhiko Yamamoto to assess this thesis.