Behavioral/Systems/Cognitive

Correlated Coding of Motivation and Outcome of Decision by Dopamine Neurons

Takemasa Satoh, Sadamu Nakai, Tatsuo Sato, and Minoru Kimura

Department of Physiology, Kyoto Prefectural University of Medicine, Kawaramachi-Hirokoji, Kamigyo-ku, Kyoto 602-8566, Japan

We recorded the activity of midbrain dopamine neurons in an instrumental conditioning task in which monkeys made a series of behavioral decisions on the basis of distinct reward expectations. Dopamine neurons responded to the first visual cue that 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 the CS was approximately proportional to reward expectations but with considerable discrepancy. In contrast, CS responses appear to represent motivational properties, because their magnitude at trials with identical reward expectation had significant negative correlation with reaction times of the animal after the CS. Dopamine neurons also responded to reinforcers that 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, whereas coding of motivational properties remained consistent during the learning. We found that the magnitude of CS responses was positively correlated with that of reinforcers, suggesting a modulation of the effectiveness of REEs as a teaching signal by motivation. For instance, rate of learning could be faster when animals are motivated, whereas it could be slower when less motivated, even at identical REEs. Therefore, the dual correlated coding of motivation and REEs suggested the involvement of the dopamine system, both in reinforcement in more elaborate ways than currently proposed and in motivational function in reward-based decision-making and learning.

Key words: **dopamine neuron; learning; teaching signal; reward prediction error; trial and error; decision-making**

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 substrates for 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 the initial phase of learning but respond to CS associated with reward in the 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 an expected reward is withheld, and maintain a baseline discharge rate when a reward is retrieved as expected (Schultz et al., 1997; Schultz, 1998). These observations led them to propose a hypothesis in which reward expectation errors (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).

In contrast, a substantial body of evidence suggests the involvement of DA systems in the processes of motivation (Robbins and Everitt, 1996; Koepp et al., 1998; Salamone and Correa, 2002; Wise, 2002), switching attention 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 who made a series of behavioral decisions on the basis of trial-specific reward expectations. Neuronal responses to CS were approximately proportional to reward expectations but with considerable discrepancy. However, the CS responses appear to represent motiva-

Received June 16, 2003; revised Sept. 8, 2003; accepted Sept. 9, 2003.

This study was supported by a Grant-in-Aid for Scientific Research on Priority Areas (C)–Advanced Brain Science Project (B) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (M.K.). We thank Dr. Keisuke Toyama, Dr. Kenji Doya, Kazuyuki Samejima, Peter Dayan, and Yael Niv for valuable comments regarding the original version of this manuscript, Dr. Yoshio Imahori for assistance in performing the surgery, and Harue Matsuda and Ryoko Sakane for technical assistance.

Correspondence should be addressed to Minoru Kimura, Department of Physiology, Kyoto Prefectural University of Medicine, Kawaramachi-Hirokoji, Kamigyo-ku, Kyoto 602-8566, Japan. E-mail: mkimura@basic.kpu-m.ac.jp.

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[a] (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[b] (44 DA cells)	18.3 ± 5.3	51.5 ± 10.0	89.9 ± 8.6	99.0 ± 2.1	98.3 ± 2.4
Early plus 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[c] (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[d] (37 DA cells)	16.8 ± 3.4	48.9 ± 8.4	92.1 ± 8.6	99.1 ± 1.4	99.3 ± 2.0
Early plus 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 ± SD of percentages.

[a] Early stage of learning, days 1-36 of the study.

[b] Late stage of learning, day 37 to month 7.

[c] Early stage of learning, days 1-15 of the study.

[d] Late stage of learning, day 16 to month 3.

tional properties, because their magnitude at trials with identical reward expectation had significant negative correlation with reaction times (RTs) of the animal after the CS. The responses to reinforcer stimuli that occurred after the behavioral decisions (outcomes) precisely encoded positive and negative REEs. The magnitude of responses to CS was positively correlated with that of outcome, suggesting modulation of REE coding by motivation. The dual correlated coding of motivation and REEs suggested the involvement of the 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's *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 hr, i.m.) was given as needed. The recording chamber was positioned at an angle of 45° 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 placed 27 cm in front of their faces. A small rectangular push button with red light-emitting diode (LED) (start LED, 14 × 14 mm) at the bottom, three push buttons with green LEDs (target LEDs, 14 × 14 mm) in the middle row, and a small red LED (GO LED, 4 mm diameter) just above the center push buttons (see Fig. 1*A*) were on the panel. 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 msec after the monkeys continued to hold the button. The target LEDs and a GO LED were then simultaneously turned on. The monkeys were required to continue depressing the start button for variable lengths of time between 0.6 and 0.8 sec before the GO LED was turned off. They released the start button and depressed one of three illuminated target buttons. If an incorrect button was depressed, a beep sound (BEEP) with a low tone (300 Hz, 100 msec) occurred with a delay of 500 msec, and the next trial began by illuminating the start LED at 7.5 sec after releasing the depressed button. Because the monkey remembered the incorrect button selected at the first trial, a choice was then made 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, 100 msec) occurred with a delay of 500 msec, 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 flashed at the same time for 100 msec to inform the animal of the end of a block of trials. At 3.5 sec 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 (see Fig. 1*B*). The first epoch was the trial-and-error epoch, in which the monkey searched for the correct button on a trial-and-error basis. Three types of trials occurred. The first was 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 found the correct button during the trial-and-error epoch. Two types of trials occurred: the first and 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 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 five 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; Frederick Haer Company, Bowdoinham, ME) 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. Bandpass filters (50 Hz to 3 kHz bandpass with a 6 dB per octave roll-off) were used. The action potentials of single neurons were isolated by using a spike sorter with a template-matching algorithm (MSD4; αOmega; Nazare), and the duration of negative-going spikes was determined at a resolution of 0.04 msec. 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, Japan). We searched for dopamine neurons in and around the pars compacta of the substantia nigra (SNc). Electrode penetrations at a 45° angle through the posterior putamen, the external and internal globus pallidus, and the 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 on the basis of the following four criteria. First, the action potentials of dopamine neurons have a relatively long duration (range, 1.6–2.9 msec; 2.2 ± 0.3 msec; mean ± SD) (see Fig. 3*A*). Second, the background discharge rate of the dopamine neurons is low (0.5–7.4 impulses per second; 4.0 ± 1.6 impulses per second) and in sharp contrast to the high background discharge rate of neurons in the substantia nigra pars reticulata (SNr). Third, under the histological reconstruction of electrode tracks in relationship to electrolytic lesion marks (a total of six marks made by passing a positive DC of 25 μA for 30 sec), the recording sites were located in the SNc or ventral tegmental area (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, multistranded, Teflon-coated, stainless-steel wire electrodes (AS631; Cooner Wire, Chatsworth, CA) 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 a small number of experiments (10 recording days in monkey DN, 5 d 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, Japan) 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 approximately ±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 (%) and reward expectation error (%) in the five 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] × volume of reward (ml) × 100, and

Negative REEs (%) = [occurrence of no reward (0) − probability of reward] × volume of reward (ml) × 100.

The responses of neurons were determined in perievent-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 msec 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, monkey DN was anesthetized with an overdose of pentobarbital sodium (90 mg/kg, i.m.) and perfused with 4% paraformaldehyde in 0.1 M phosphate buffer through the left ventricle. 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 on the basis of six electrolytic microlesions (Fig. 3*B*,*C*). Sections spaced at 300 μm intervals through the striatum and substantia nigra were stained for tyrosine hydroxylase (TH)-like immunoreactivity (anti-TH; 1:1000; Chemicon, Temecula, CA) (Matsumoto et al., 1999).

Results

Evolution of task performance through learning act-outcome relationships

The monkeys chose one of three potentially correct buttons as their first choice (Fig. 1*A*,*B*) (trial-and-error epoch; see Materials and Methods). They received 0.35 ml of reward water if their choice was correct. However, if their choice was incorrect, they made a second choice from the remaining two buttons. If their second choice was incorrect, they chose the one remaining button and received a reward in all trials, except for ~10% of the trials in which monkeys made errors again. Thus, there were three 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 button that was previously found correct for two succeeding trials (Fig. 1*B*) (repetition epoch, R1 and R2 trials). Figure 1*C* plots the average correct choice rates in five trial types over the entire recording period in monkey DN.

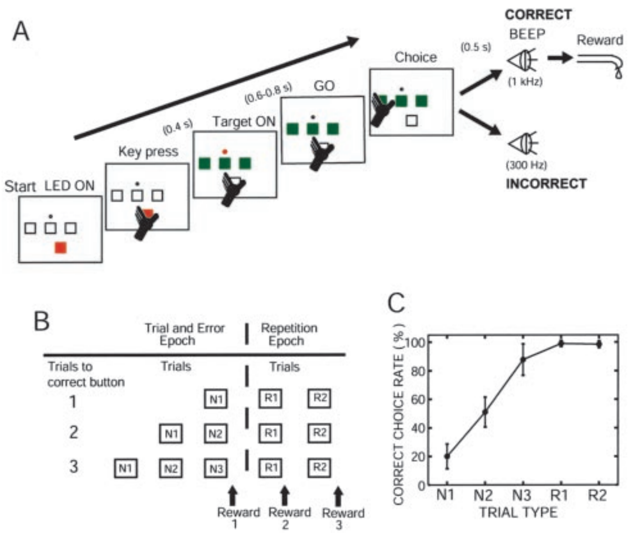

Figure 1. Behavioral task, trial types, and percentage of correctness at each trial type. *A*, Illustration of sensorimotor events that appeared during a single trial (see details in Materials and Methods). *B*, Two epochs (trial-and-error and repetition epochs) and five trial types (N1, N2, N3, R1, R2) in a block of trials classified on the basis 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.

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 approximately one of three, 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 (one of two) and N3 (one of one) trials. In this study, it was critical that monkeys had a wide range of trial-specific reward expectations before making behavioral decisions. The average correct choice rates in the five trial types at the initial partially learned stage and at the later fully learned stage are shown in Figure 2*A* and 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 correct in the previous set during the N1 trials in both monkey DN (average, 62%) and monkey SK (average, 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 a conditioned stimulus, with respect to the

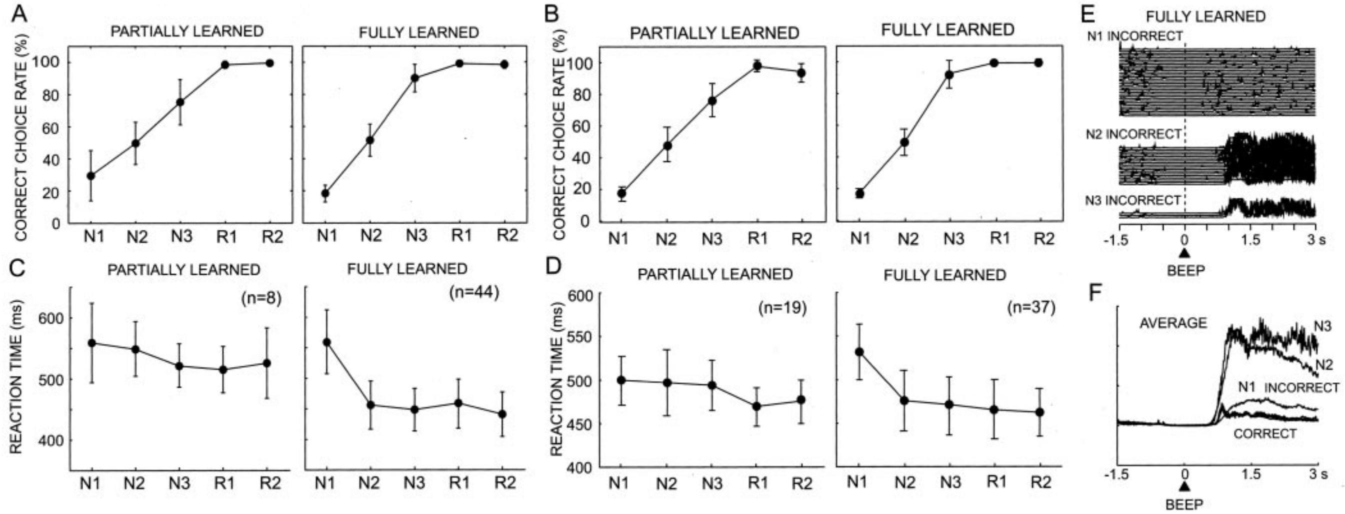

Figure 2. Task performance in the partially learned and fully learned stages. *A*, Correct choice rate against the trial types in partially learned (1–36 d) and fully learned (37–215 d) stages in monkey DN. *B*, Same as *A* but in the partially learned (1–15 d) and fully learned (16 –95 d) stages for 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 three incorrect trial types (N1, N2, N3) (left) and average traces during five correct (N1, N2, N3, R1, R2) and three incorrect (N1, N2, N3) trial types (right) in monkey DN. BEEP indicates the onset of the beep sound after the animal's choices.

unconditioned stimulus (reward) in the present instrumental conditioning task. The average RTs of button pressing after the presentation of a CS in all five trial types were relatively prolonged during the initial stage of learning (d 1–38, 533.3 ± 49.7 msec in monkey DN; d 1–15, 486.6 ± 31.1 msec in monkey SK; mean ± SD). There was no significant difference among the RTs in the five 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 five 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 msec in monkey DN; 468.0 ± 32.7 msec in monkey SK), whereas those of the N1 trials became longer (559.0 ± 52.9 msec in monkey DN; 531.2 ± 32.0 msec in monkey SK). The RTs in the N1 trials were significantly longer than those in the other four 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). On the basis of these learning stage-dependent differences in task performance, the experimental sessions were separated into an early stage, a partially learned stage, and a later fully learned stage. The average RTs at the two stages in the two monkeys, thus defined, are plotted in Figure 2, *C* and *D*.

Monkey DN developed a characteristic orofacial reaction after incorrect trials at the fully learned stage. The electromyograms of digastric muscle activity during the consumption of liquid reward revealed similar activity patterns for all five types of correct trials (Fig. 2*E*). In contrast, during the incorrect trials, 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*). The activity in the N1 trials was smallest, in which the reward probability was lowest, although it was slightly larger than that in the correct trials. The muscle activation probably reflected levels of the 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 the 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, and thus the expectation of reward, and the levels of motivation specific to each trial type through learning act-outcome relationships 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, 90 in monkey SK) in the SNc and VTA while the monkeys made a series of reward-based decisions. These neurons had characteristic discharge properties that were 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 four impulses per second (see Materials and Methods). The properties of these DA neuron discharges significantly differed from those of neurons in the nearby SNr, as shown in Figure 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 of 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 >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*,*C*). DA neurons recorded in the SNc and VTA of monkey DN are 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 and then entered 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.

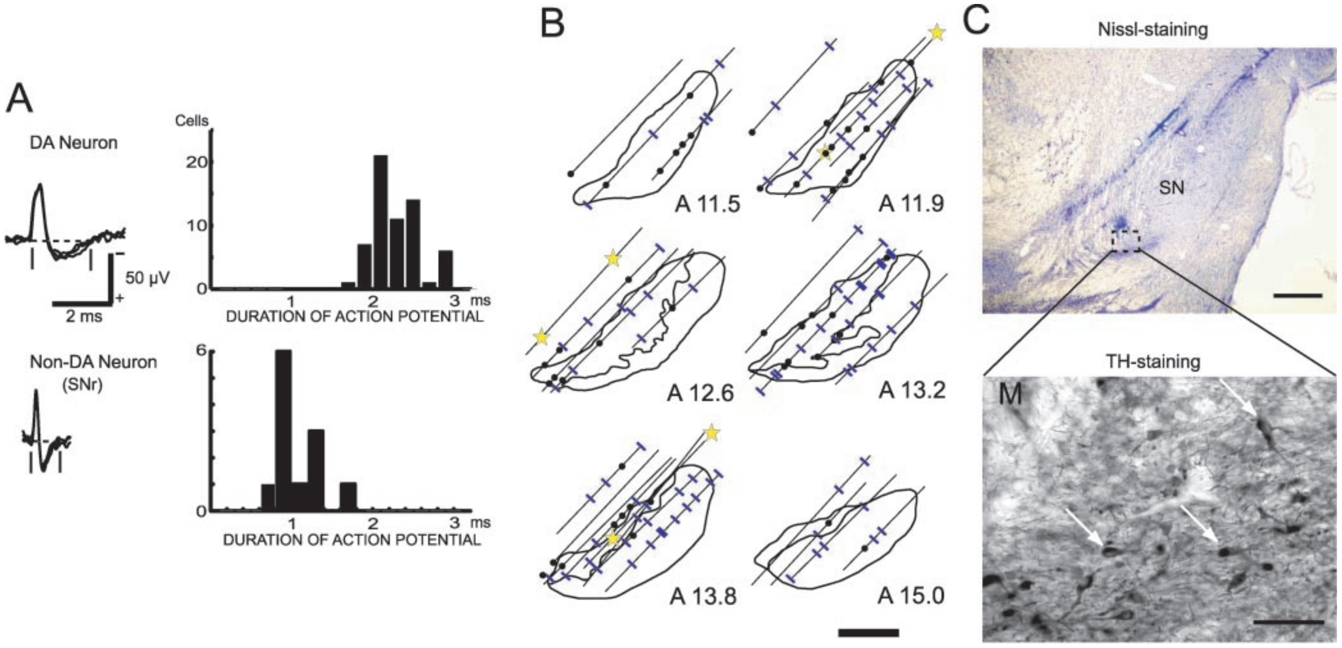

Figure 3. Electrophysiological and histological identification of DA neurons. *A*, Left, Superimposed traces of extracellularly recorded action potentials of DA neurons (SNc) and non-DA neurons (SNr). The two vertical lines and the horizontal interrupted line indicate how the duration of the action potential was measured. Right, Histograms of the duration of recorded action potentials. *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 microlesion marks. Scale bar, 2 mm. *C*, A Nissl-stained section at the level of the SN is shown (scale bar, 1 mm) (top), and part (interrupted circle) of the neighboring TH-stained section is shown at higher magnification (scale bar, 100 μm) (bottom). White arrows, TH-immunoreactive neurons; M, part of a lesion mark.

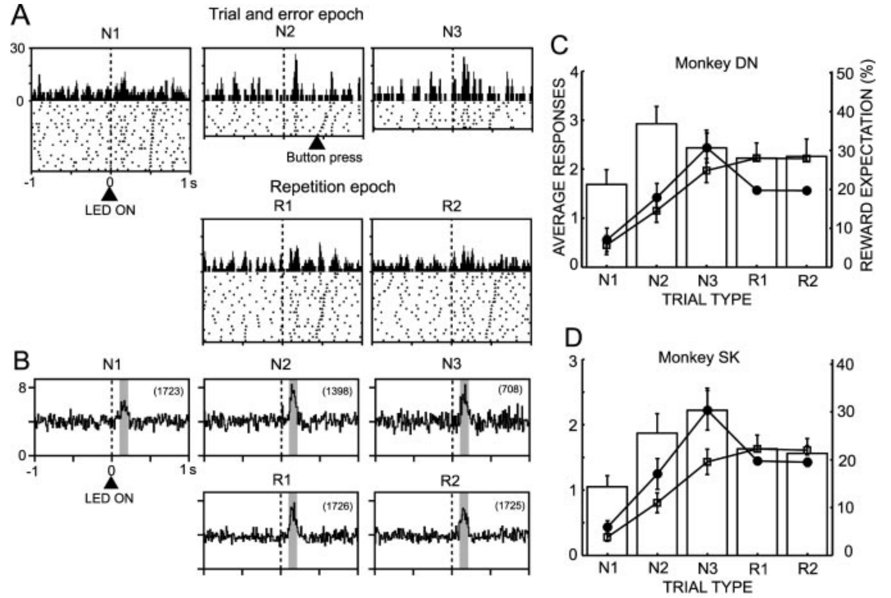

Figure 4. Response of DA neurons to the start LED (CS). *A*, Activity of a single DA neuron recorded in the SNc of monkey DN before and after the CS in the five trial types. Impulse discharges that occurred during the individual trial types are represented separately as rasters and histograms. The activity is centered at the onset of the CS (vertical interrupted line). The trials in the raster display were reordered on the basis of the time interval between onset of the 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 the CS in monkey DN. *C*, Average increase in the discharge rate of 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 msec period just preceding the onset of the CS. The results are shown as means ± SE 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) (see Results for explanation). The 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 msec.

Therefore, the activity of 56 presumed DA neurons identified on this basis is described as a separate neuronal population for monkey SK from that of monkey DN.

Responses to CS in instrumental conditioning

The DA neurons either increased or decreased their tonic discharges after two different sensory events occurred in the task. The first was the CS, which instructed the monkeys to initiate each trial of the instrumental task. The second was a high-tone or low-tone BEEP reporting that either the animal's choices were correct and the reward would come or the choices were incorrect and no reward would be given. The high-tone and low-tone BEEPs 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 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 in ensemble average responses in Figure 4, *A* and *B*. In Figure 4, *C* and *D* are

Table 2. Number of responsive DA neurons to the CS and reinforcers

	Monkey DN		Monkey SK	
	Partially learned stage	Fully learned stage	Partially learned stage	Fully learned stage
CS	4	23	11	16
Reinforcers	2	30	12	26
CS and reinforcers	1	18	7	16
Total	8	44	19	37

Numbers represent responsive neurons determined by Wilcoxon single-rank test at $p < 0.05$ (Kimura, 1986).

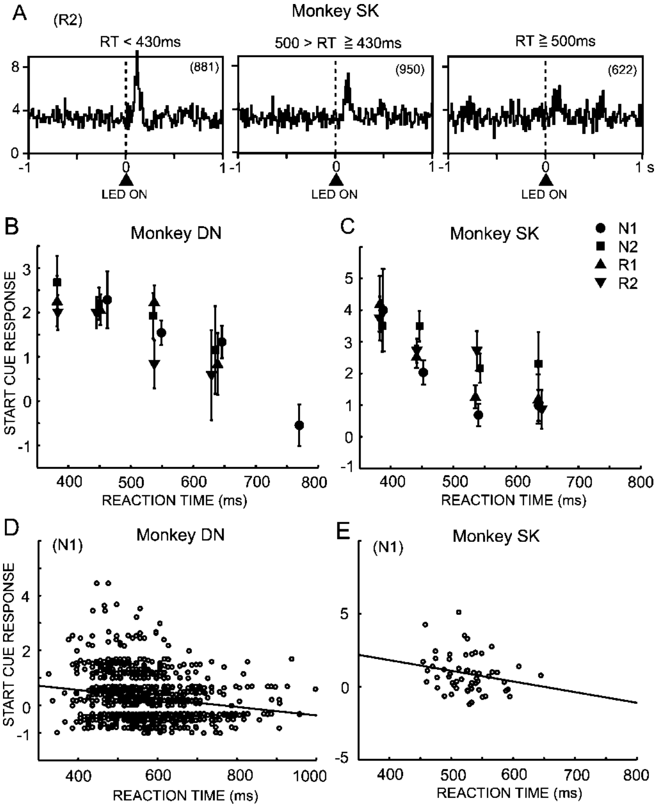

Figure 5. Relationship of response magnitudes of DA neurons to briskness of behavioral responses to the CS. *A*, Population response histograms of the 56 DA neurons in monkey SK to the CS during R2 trials. Histograms are separated on the basis of the trials with short, middle, and long RTs to CS. The number in parentheses indicates the number of trials involved in each histogram. *B*, Correlation of the magnitude of neural responses to the CS in 52 DA neurons in monkey DN to RTs to depress the start button after the CS. The correlations are plotted separately in N1, N2, R1, and R2 trials. The results of N3 trials are not plotted because of the very small number of trials. The trials were classified into five groups on the basis of 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 on the basis of the RTs of trials during recording of 56 DA neurons. Because the RTs in monkey SK were shorter than those in monkey DN by an average of ~80 msec, the ranges of RTs in three groups of trials in monkey SK were shifted to shorter RTs from those in monkey DN. *D*, Correlation of the magnitude of responses to the CS with RTs in each trial in monkey DN. The correlation analysis was performed on 854 trials from 27 neurons showing significant responses to the CS. *E*, Correlation between average CS responses of single neurons and average RTs in monkey SK (56 neurons).

plotted against trial type average increases or decreases of discharges from the baseline level in response to CS in two monkeys. Approximately one-half of neurons showed significant responses to CS (Table 2) (27 of 52 neurons in monkey DN; 27 of 56 neurons in monkey SK). The results are based on the neuronal activity in both the partially learned and fully learned stages. The average response was an increase in the discharges in all trial types. In monkey SK, the responses in N1 trials were the smallest among the five 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 the animal's expectation of reward, because it is supposed in the reinforcement learning algorithm that the responses to the CS represent a 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*,*D*). The neural responses and 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 reward expectations as the probability of reward (Fig. 4*C*,*D*, 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 five trial types. The reward expectations as the product of probability and volume of reward (Fig. 4*C*,*D*, filled circles) did not estimate the responses very well, 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 an animal's motivation to work for a reward. We used a time for monkeys to depress the start button (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 the 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 the amplitude of DA neuron response within single-trial types in which monkeys performed the trial with a consistent level of reward expectations. Figure 5*A* shows ensemble averages of neuronal activity of the three groups of R2 trials with short, middle, and long RTs in monkey SK. The largest activation occurred in the short RT trial group, the smallest activation occurred at the longest RT group, and the middle level of activation occurred in the middle RT group. Figure 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 (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 the RT 500–600 msec 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 of 52 neurons in monkey DN and 12 of 56 neurons in monkey SK.

In most N1 trials, the monkeys chose the button that had been correct in the previous set (average: 52% in monkey DN, 98% in monkey SK). There was no significant influence of the tendency on the neuronal responses to the CS at N1 trials ($p > 0.3$ for monkey DN; Wilcoxon rank test). The measurement of eye position signals during task performance revealed that monkeys

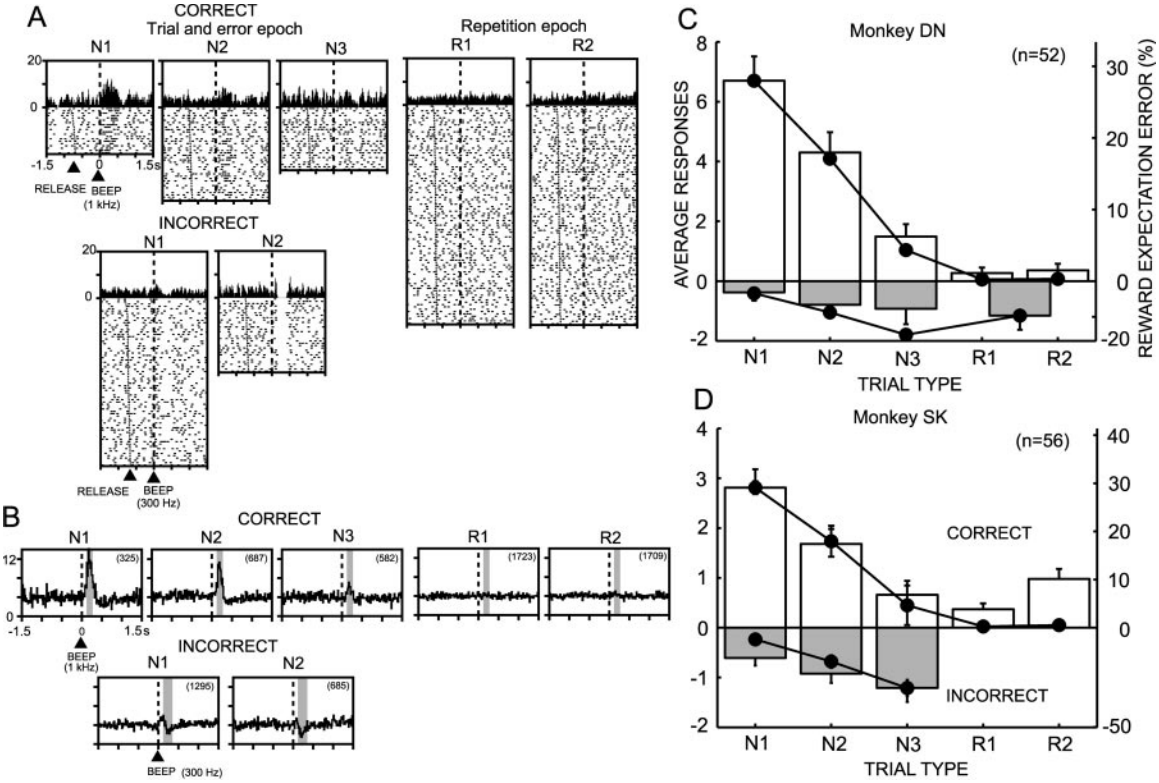

Figure 6. Responses of DA neurons to reinforcers after the animal's choices at each task trial. *A*, Activity of a representative DA neuron at correct and incorrect choices in the five trial types. The displays are centered at the onset of the reinforcers (vertical interrupted lines). The trials in the 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 five trial types. The number 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 msec period just preceding the onset of the 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 Materials and Methods). *D*, Same as *C* but for monkey SK. Because incorrect trials rarely occurred during the repetition epoch, the neuronal responses and REEs for R1 and R2 trials were either combined and plotted as a single-trial type in monkey DN or not shown in monkey SK.

were looking at either one of three target buttons or a hold button before illumination of the hold button (CS) in most of the time. Specifically, during 500 msec before the CS appearance, monkeys tended to look at the 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 a variance in the RT of depressing the hold button. However, limited amount of eye movement data in the present study did not allow us to draw definitive conclusions on this issue. This is an 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 (during the initial 3 hr) of the daily experimental schedule when the monkeys were thirsty and those performed during the later session (after the initial 3 hr) 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 the later session by 21 ± 5.3 msec (mean ± SD in five trial types; $p < 0.001$ in N2, N3, R1, and R2; Mann–Whitney *U* test) in monkey DN, and by 10.6 ± 12.5 msec 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 on other days of the week by 25.9 ± 22.7 msec (mean ± SD in five 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 they were less motivated to work on Monday. Thus, these results support the motivation hypothesis.

Coding outcomes of behavioral decision

DA neurons characteristically responded to the reinforcers after behavioral decisions. Figure 6*A* illustrates representative responses of a DA neuron in the SNc of monkey DN to the positive reinforcer after correct choices (all five 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 discharge rate (positive response). In contrast, the negative reinforcer produced the decrease of discharges; the negative response was preceded in many cases by a small transient increase of discharges. The population response histograms of the 52 neurons in monkey DN in Figure 6*B* demonstrated a systematic dependency of neuronal responses on the trial type. These relationships 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 Figure 6*C* and *D*. In both monkeys, the positive responses were highest at N1, became smaller at N2, and became smaller still at N3 trials. There was nearly no response at the repetition epoch (R1 and R2 trials) in monkey DN (Fig. 6*C*), whereas there were small responses in monkey SK (Fig. 6*D*). More than 60% of neurons responded to the reinforcers (Table 2) (32 of 52 in monkey DN; 39 of 56 in monkey SK). The recording sites of neurons responsive only to start cue, only to reinforcers, and to both start cue and reinforcers in monkey DN were histologically reconstructed in the midbrain (see supplementary data; available at www.jneurosci.org). However, 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 REEs. Positive and negative REEs derived from the product of probability and volume of reward at each trial type (see Materials and Methods) are superimposed on the response histograms in Figure 6*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 per percentage of REEs in monkey DN and 0.026 impulses per percentage of REEs in monkey SK. There was a weaker correlation between the negative responses and REEs in monkey DN ($r = 0.087$; gain, 0.007 impulses per percentage of REEs) and in monkey SK ($r = 0.159$; $p < 0.05$; gain, 0.009 impulses per percentage of REEs).

The positive responses at N1 trials after monkey DN chose the button that had been correct in the previous set (average, 7.11 ± 0.98 impulses per second) were slightly larger than those when a previously incorrect button was chosen (average, 5.46 ± 0.90 impulses per second; $p < 0.05$; 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 a negative reinforcer also encode REEs, although the gain in encoding by decreasing the discharge rate is smaller than that for the positive reinforcer.

Positive relationship 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 relationship of responses to CS and those to positive reinforcers (high-tone beep) in a 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 the CS were positively correlated with those to positive reinforcers in monkey DN (Fig. 7*A*) ($r = 0.234$; 52 neurons) and in monkey SK (Fig. 7*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. Thus, the results support an interesting view that the number of DA neuron spikes encoding REEs, a gain of coding REEs, might be positively modulated by the responses to the CS that appear to reflect levels of motivation.

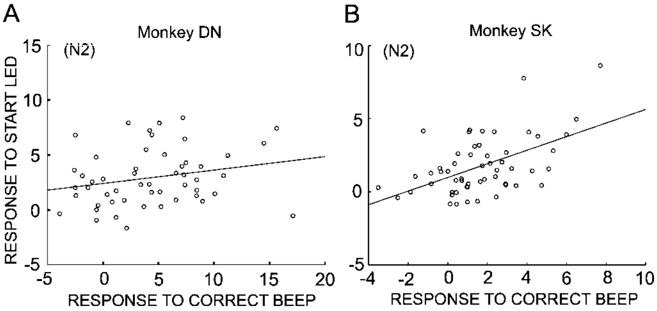

Figure 7. Relationship of responses to the CS and response to high-tone beep, positive reinforcer. *A*, Scatter plot showing positive correlation between the response to the CS and response to positive reinforcer in N2 trials in monkey DN ($r = 0.234$; slope = 0.125). *B*, Same as *A* but for monkey SK ($r = 0.524$; slope = 0.551).

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 8*A* plots responses to the 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, the slope of the correlation was consistently maintained in the two stages of learning in the two monkeys, although the correlation in the partially learned stage of monkey DN was not significant, probably because a small number of neurons was 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 that occurred after the 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. 8*B*, left). In contrast, in the fully learned stage when RTs after the start cue at N1 trials became significantly longer than those at the other four trial types because of very low reward expectation, much stronger positive responses appeared in correct N1 and N2 trials (Fig. 8*B*, right). An approximately fourfold increase occurred in the gain of coding positive REEs in monkey DN. A similar but mild increase was also observed in monkey SK. However, 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. 8*A*). These observations indicated that the coding of REEs by DA neuron activity develops through the process of learning act-outcome relationships in the reward-based decision-making task, whereas motivational properties attributed to the CS appear at an initial stage of learning and are maintained during learning.

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 the CS appear to represent motivational properties attributed to the CS. Second, the responses of DA neurons to a positive reinforcer after behavioral decisions precisely encode REEs. The responses to a negative reinforcer also encode REEs, but the gain of encoding by decreasing the discharge rate is

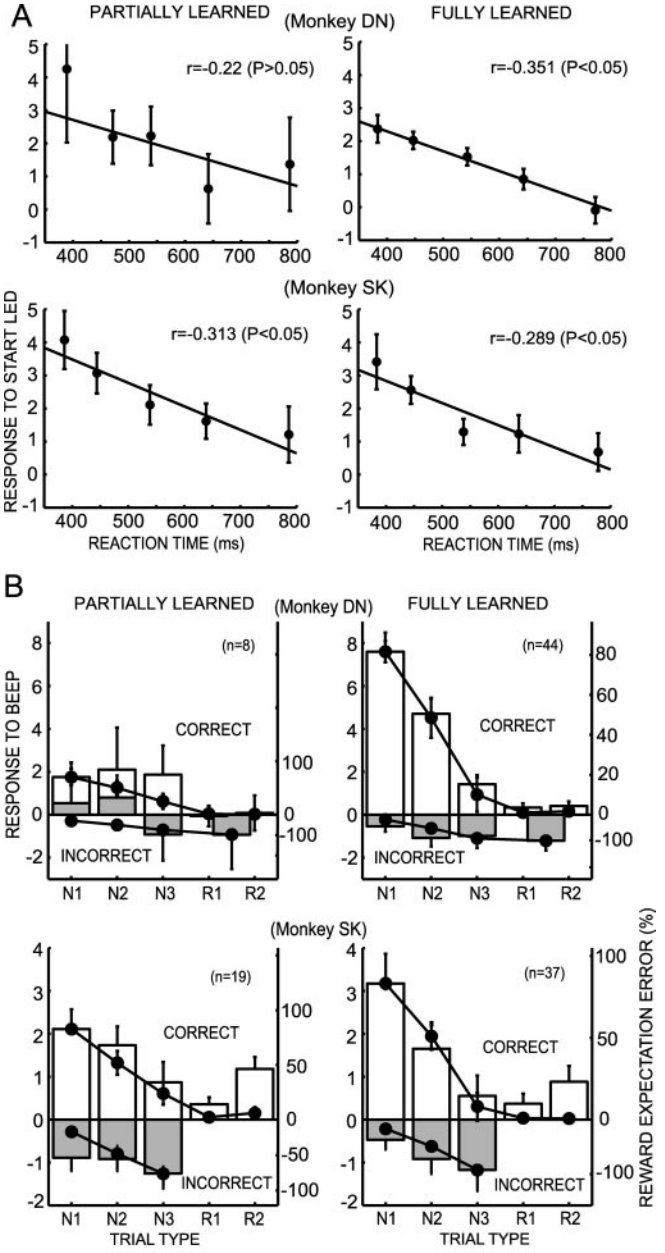

Figure 8. Responses of DA neurons at the early partially learned stage and the later fully learned stage. *A*, Scatter plot of the average responses of DA neurons (mean and SEM) and RTs to depress the start button after the CS. The plots were made for all trials independent of trial type. Trials were divided into five groups on the basis of 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 five 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 to have the same value at the maximum REE.

much smaller than that for a positive reinforcer. This finding is in agreement with that of Fiorillo et al. (2003), in which phasic responses of DA neurons varied monotonically across the full range of reward probabilities. However, 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 the 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 relationships through a remarkable increase of gain of coding, whereas coding motivational attribution to the CS appears at an initial stage of learning and is consistently maintained through the entire learning process.

Dual and correlated coding of motivation and 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 in which behavior is directly strengthened or weakened by the consequence that follows (Thorndike, 1911). Reinforcement learning theories proposed a computational algorithm of reward learning in which the agent adapts its behaviors on the basis of 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 does the Pavlovian association between the stimulus and a conditioned response occur but also a motivational transformation. That is, the stimulus takes on specific motivational incentive properties (CS) originally possessed 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 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 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 an increase of DA neuron spikes, and thus an 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 a 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, similar to 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 stay with the old one (wait for the next start signal). However, 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 the 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 the CS while still having considerable discrepancy. In contrast, we found, for the first time, that the responses to the CS were correlated significantly with RTs at trials with an identical level of reward expectation. Thus, it appears that the responses to the 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 the 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 PPTN to DA neurons in SNc have been demonstrated previously (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 approximately one-half of DA neurons studied significantly responded to both CS and reinforcers (Table 2) (19 of 52 in monkey DN; 23 of 56 in monkey SK). What is the functional significance of the dual coding of incentive attribution to the 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 a teaching signal by a motivation. For instance, the 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, whereas it is slower when less motivated, even at identical REEs as a consequence of an action. In the present study, there was a positive correlation between the responses to the CS and those to positive reinforcers (Fig. 7). This suggests a new and richer model for DA neurons as teaching signals in reinforcement learning than currently proposed. In contrast, this is consistent with the theory of classical conditioning in which the 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 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).

The principal functions of reward are supposed to produce satisfaction, elicit approach behaviors, and reinforce immediately preceding actions (Thorndike, 1911; Hull, 1943; Olds and Milner, 1954). The DA neuron responses to the CS may participate in producing satisfaction and eliciting approach behaviors, whereas 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, whereas relatively little attention has been shown to the responses to the CS and to their functional significance. However, 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 the 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 the reinforcement process. This view was supported by the observation that the responses to the CS and 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 the 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 and thus DA release during learning. A critical question arises: Why were the target stimuli or GO signals ineffective in eliciting DA neuron responses? An analysis of eye movements revealed a tendency of monkeys to make saccade before and after the CS frequently to one of the targets that was going to be chosen after GO signal, and suggested that the monkeys made a decision close to the CS with trial type-dependent reward expectations. These observations could support the view that the DA responses to the CS motivate the entire trial. An 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 the CS is also involved, because attention can contribute to shaping new forms of behaviors toward the direction of their goal (i.e., the reward) (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 (Dickoff J, James P, translators). Indianapolis: Bobbs-Merrill.

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, MA: MIT.

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

Bolles RC (1972) Reinforcement, expectancy, and learning. Psychol Rev 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 [Suppl] 3:1218–1223.

Dickinson A (1980) Contemporary animal learning theory. Cambridge, UK: Cambridge UP.

Dickinson A, Balleine B (1994) Motivational control of goal-directed action. Anim Learn Behav 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 (Houk JC, Davis JL, Beiser DG, eds), pp 249–270. Cambridge, MA: MIT.

Hull CL (1943) Principles of behavior, an introduction to behavior theory, Chap 6, 7. New York: Appleton-Century.

Jacobs BL (1986) Single unit activity of brain monoamine-containing neurons in freely moving animals. Ann NY 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: Chicago UP.

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.

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 (Black AH, Prokasy WF, eds), 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, pp 87–160. Cambridge, MA: MIT.

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: Macmillan.

Toates F (1986) Motivational systems. Cambridge, UK: Cambridge UP.

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.