Supporting Information

Enomoto et al. 10.1073/pnas.1014457108

SI Methods

Animals and Surgery. Three Japanese monkeys (*Macaca fuscata*; monkey BT, male, 8.3 kg; monkey SK, female, 8.1 kg; and monkey CC, female, 7.5 kg) were used in this study. Four head-restraining bolts and two stainless-steel recording chambers were implanted on their skulls by using standard surgical procedures. The monkeys were anesthetized with ketamine hydrochloride (10 mg/kg, i.m.) and sodium pentobarbital (Nembutal; 27.5 mg/kg, i.p.). Two recording chambers were laterally positioned under stereotaxic guidance at an angle of 45° to record the activity of dopamine neurons from the midbrain in both hemispheres. The center of each chamber was adjusted according to Horsley–Clark stereotaxic coordinates: lateral, 15 mm; anterior, 12 mm from the interaural line and 12 mm below the skull surface.

Multistep Choice Task. The monkeys sat in a primate chair facing a task panel that was placed 25 cm in front of their faces. On the panel was a small, rectangular push button with a red LED (start cue, 14 × 14 mm) at the bottom, three push buttons with red LEDs (target LEDs, 14 × 14 mm) in the middle row, and a small, red LED (go LED, 4 mm diameter) just above the center push buttons (Fig. 1*A*). Each trial of the multistep choice task was initiated by illumination of the push button start cue. The monkeys depressed the illuminated start button with the hand contralateral to the neuronal recording. After they depressed the start button for 400 ms, the three target LEDs and a go LED were simultaneously turned on. The monkeys depressed the start button for another 700 ms until the go LED was turned off. They released the start button and depressed one of the three illuminated target buttons (N1 trials). If a correct button was depressed, a beep sound with a high tone (1 kHz for 100 ms) sounded with a delay of 300 ms as a positive reinforcer, and a small amount of reward water (0.35 mL) was delivered through the spout attached to the monkey's mouth. If an incorrect button was chosen, a low-toned beep (300 Hz for 100 ms) occurred with a delay of 300 ms as a negative reinforcer, and no reward was given. The next trial began by the illumination of the start cue at various time intervals (6.5–8.5 s) after the monkeys had released the target button. All three targets were once more illuminated, and the monkeys chose again (N2 trials) with the same consequences as earlier. If they made another incorrect choice, the third trial started, and they again chose among the three targets (N3 trials). The rewarded target during the exploration trials was also rewarding during the following exploitation trials. Therefore, in the ideal situation, the monkeys correctly discovered the rewarding target during the exploration trials and exploited this knowledge during the next two trials, receiving reward water three times through the choices of the same target button (Fig. 1*B*): for the first step of the exploration epoch (N1, N2, and N3 trials) and for the second and third steps during the exploitation epoch (R1 and R2 trials). To let the monkeys know that the three-step choice activity was complete, all LEDs were flashed simultaneously for 100 ms, 2 s after the monkeys released the target button following the correct R2 trials. At 4.5 to 6.5 s after the flashing, the next three-step choice action began with a new correct target at an unpredictable location. In the two-step choice task, the monkey received two rewards (no R2 trial). The correct choice rate of the N1 trials was less than chance, one-third, and approximately 20% under computer control (Fig. 1*C*). Monkey SK received smaller volumes for the second and third rewards (0.2 mL) than that for the first reward (0.35 mL) to quantitatively evaluate a representation of multiple future rewards.

Data Recording. We used conventional electrophysiological methods for single-unit recording with epoxy-coated tungsten microelectrodes (FHC) (1). We searched for dopamine neurons in and around the pars compacta of the SNc and VTA. The electrodes were inserted by an oil-driven micromanipulator (MO-97-S; Narishige) at a 45° angle through the posterior putamen, the external and internal globus pallidus, and the internal capsule before reaching the midbrain. The characteristic depth profiles of the electrical activities in these structures facilitated our approach to the dopamine neurons. The neuronal activity was amplified and displayed on an oscilloscope by using conventional electrophysiological techniques. Band-pass filters (50 Hz to 1 kHz) were used to tune the amplifier system to sample neural action potentials with low noise levels. The action potentials of single neurons were isolated using a spike sorter with a template-matching algorithm (MSD; Alpha Omega Technologies), and the onset times of the action potentials were recorded on a laboratory, custom-made DOS/V computer with the onset and offset times of the stimuli and the behavioral events that occurred in association with the tasks. In accordance with previous studies (1–3), we identified dopamine neurons based on the following four criteria: (*i*) an action potential with long duration (>1.5 ms, 2.2 ± 0.3 ms, mean ± SD), (*ii*) a low background discharge rate (4.0 ± 1.4 spikes/s), (*iii*) a phasic increase in discharge rate following the unexpected delivery of reward water, and (*iv*) histological verification (Fig. S1). Neuronal recordings were performed during both the early learning stage of the three-step choice task in monkey SK and during the late advanced stages of the two- or three-step choice tasks in all monkeys (Table S1). Orofacial movements in anticipation of reward water were monitored and recorded at 10-ms intervals by using a strain gauge (KFG-2N amplified by DPM-711B; Kyowa) attached to the spout in monkeys BT and CC.

Data Analysis. We identified dopamine neurons and isolated their action potentials in more than 130 trials (at least 10 N3 trials per neuron) during multistep choice tasks and 150 trials during classical conditioning tasks (>25 trials in all trial types per neuron). Spike density histograms of dopamine neuronal activities were constructed in relation to task events. Neuronal activity was regarded as task-related if the discharge rates after either the task start cue or the reinforcer beeps during multistep choice tasks or after the CSs during classical conditioning increased or decreased significantly from baseline discharge levels recorded during the 500- to 750-ms test window (25 bins) before the presentation of the task start cue. Test windows were shifted in 10-ms bins up to 450 ms starting from the onset of an event. The activity was considered significant if more than three consecutive comparisons between the test and the baseline windows were significantly different (two-tailed Wilcoxon test at $P < 0.05$) (4). The onset and offset of a response were taken to be the beginning and end, respectively, of the significant changes in activity. Quantification windows for the measurement of the magnitude of the dopamine neuronal activity were set one SD wider than the windows determined by the average onset and offset times of significant changes in discharge rate in all examined neurons.

We analyzed the licking data in all of the sessions that contained more than 130 trials during the multistep choice tasks and all neuronal activity recording sessions during classical conditioning tasks. The duration of the anticipatory licking movements was defined as the total time during which the amplitude of the licking movements exceeded the threshold (50% and 30% maximum in multistep and classical conditioning tasks, respectively). The duration was quantified in a 50-ms time window, which was shifted in 10-ms steps from 100 to 800 ms before the appearance of reinforcer beeps in the multistep task and in the 700 ms period just before the beep sound in the classical conditioning paradigm. Mean anticipatory licking durations (ms/trial) were normalized to the value in the trial type that showed the maximum duration.

Neural activities and anticipatory licking durations in each trial type were compared by using multiple, two-sample comparisons corrected by the Tukey–Kramer method. For the multistep choice task, normalized average anticipatory licking durations, neuronal discharge rates after the start cues, and neuronal discharge rates after reinforcers for correct and incorrect choices were quantitatively examined based on a model of reinforcement learning theory, as described later.

Simulation of Duration of Anticipatory Licking and Dopamine Neuron Activity Based on Value Function of Reinforcement Learning Theory. To understand the computational mechanisms that may be used by the monkeys and how dopamine neurons encode values during a multistep choice task, we adopted the TD learning model because it is the most standard reinforcement learning algorithm. The key quantity of the TD model is the so-called value function, $\mathrm{V(S_t)}$, which represents the sum of expected future rewards ($\mathrm{r_t}$) discounted by the number of steps to obtain them, starting at state (context)$\mathrm{S_t}$, and is defined as follows:

$$V(S_t) = E\left\{\left(\sum_{k=0}^{\infty}\gamma_{t+k+1}|S_t = S\right)\right\}, \qquad \textbf{[S1]}$$

where E represents the expectation taken over all contexts (states) and k is an index for future steps. In the present experiment, the state $\mathrm{S_t}$ takes on values N1, N2, N3, R1, or R2, with R2 as the terminal state. The only adjustable parameter in the TD model is the discount factor, γ ($0 \le \gamma \le 1$), which controls how far the agent (i.e., monkey) looks ahead to evaluate a state. Here, we assumed that the monkey's long-term reward prediction can be captured through γ.

During learning, the TD model updates $\mathrm{V(s_t)}$ as follows in proportion to the TD error $\mathrm{\delta_t}$:

$$V(s_t) \leftarrow V(s_t) + \alpha\delta_t. \qquad \textbf{[S2]}$$

where α is the learning rate, and $\mathrm{\delta_t}$ is $\mathrm{r_t} + \gamma\mathrm{V(s_{t+1})} - \mathrm{V(s_t)}$. The first and second terms of the TD error represent the estimations of $\mathrm{V(s_t)}$ after receiving a reward at time t. The third term is the same estimation as before the receipt of the reward. Therefore, when the estimation of $\mathrm{V(s_t)}$ is complete, TD error δt should be 0. TD learning minimizes TD error.

In our simulation, we estimated γ *so* the value functions *V* best fit the duration of anticipatory licking and dopamine neuronal discharges. More specifically, we constructed a five-dimensional vector consisting of the experimental duration of licking or the mean firing of dopamine neurons in each state (i.e., N1, N2, N3, R1, and R2). Then, we ran the TD algorithm to learn the multistep choice task by using 250 trials, which also produced a five-dimensional vector of the simulated duration of licking or dopamine discharges (TD error) after the completion of learning. We then searched for a γ ($0 \le \gamma \le 1$) that maximized the correlation coefficient (*R*) between the experimental and simulated licking duration or neural firing vectors by using a MATLAB optimization function, "fmincon." We used a single seed of random number generation to simulate the stochastic selection of rewarded target and actions. To see how quickly *R* decreases as γ is changed from the optimal γ (Figs. 2*D*, 3*E*, and 4*D*), we used second derivatives of *R* (5–7). We also tested two different random number seeds and confirmed that both settings yielded the same results. Although *R* and mean squared error have been used in this type of analysis, we adopted an *R* that did not need normalization.

In this simulation, we set the value of α to 0.02. We varied the learning rate, α, from 0.0 to 1.0, and the initial values of $\mathrm{V}(s_t)$. These different settings affected the results only slightly, although the larger α gradually increased the mean and variance of the estimated γ values.

We also confirmed that simulating for 250 trials produced almost the same results as a simulation for 500 or more trials. For example, the simulation during the early and advanced stage for 250 trials gave a γ of 0.00 (the value of the second derivative of *R* was −1.9) and a γ of 0.31 (−2.2), respectively, and the simulation for 500 trials gave a γ of 0.00 (−2.1) and a γ of 0.34 (−3.9), respectively. Simulation for 1,000 trials did not make any significant improvement in the convergence of values.

Control Experiments for Different Sizes of Reward. Six months after the completion of neural recordings during the three-step choice task among three alternatives with different sizes of reward in monkey SK, we retrained her to perform a two-step choice task with fixed amounts of reward (0.35 mL) in all trials, and neural recording sessions were then restarted 3 mo later (Fig. 3*D*).

Control Experiment for Uncertainty of Decision and Reward Availability. Monkey BT performed an uncertainty-reduced, two-step choice task in which the target LEDs that were chosen but unrewarded (i.e., incorrect) in the previous one or two trials were not illuminated in the next trials (Fig. S4*A*). In other words, the number of illuminated targets was reduced after monkeys chose the target during the exploration trials, and only one available target was explicitly instructed in the N3 trials during the exploration epoch and the R1 trials during the exploitation epoch. This was in contrast to the original task, in which all three targets were illuminated in every trial type during the exploration and the exploitation epochs. Therefore, the uncertainties of the decision and the reward availability were very low in both the N3 and the R1 trials (entropy was near 0 bit), but the uncertainties in the original task, especially during the N3 trials, were considerably higher. In this way, we could manipulate the levels of uncertainty during the N3 trials by using the control task in addition to the original task. We recorded the activities of the same dopamine neurons during the original and control task in monkey BT, which performed the tasks in a block schedule (i.e., >130 trials of each task) in which the original and control tasks occurred in an arbitrary order.

Uncertainty of reward was estimated by the binary entropy function *H*(*p*), defined as follows (8):

$$H(p) = -p\log p - (1-p)\log(1-p) \qquad \textbf{[S3]}$$

where *p* represents reward probability. This function has a maximum (1 bit) at $P = 0.5$ and minimums (0 bit) at $P = 0$ and $P = 1$.

Classical Conditioning. On a panel in front of the monkeys, 10 small, green rectangular LEDs (14 × 14 mm) were embedded. During the trial, the monkeys depressed an illuminated start button at the bottom of the panel with the hand contralateral to the neuronal recording. After depressing the start button for 400 ms, the CS was turned on for 1 s. There were five different CSs, which were distinguished from one another by the spatial arrangements of three illuminated LEDs among 10 alternatives (Fig. S5*A*). Individual stimuli were associated with reward at varied probabilities of 0% (P0), 20% (P20), 50% (P50), 80% (P80), and 100% (P100). Each of the five CSs appeared in random order. In monkey CC, four CSs (P20, P50, P80, and P100) were used. A low-tone or high-tone beep followed the CS as a reinforcer that instructed that no reward or reward (0.35 mL), respectively, would follow 300 ms after the beep.

Histologic Examination. After the recording was completed, small electrolytic lesions were made at nine locations along six selected electrode tracts, both in the SNc and in the VTA, by passing a direct anodal current (20 μA) for 30 s through tungsten microelectrodes. Monkey BT was anesthetized with pentobarbital sodium (Nembutal; 85 mg/kg, i.p.) and transcardially perfused with 4% paraformaldehyde in 0.9% NaCl solution. Frozen sections were cut every 50 μm in the frontal plane parallel to the electrode penetrations and stained with cresyl violet. We reconstructed the electrode tracts and sites of the recorded neurons in the midbrain based on the electrolytic microlesions. To determine the identity of the recorded neurons, sections spaced at 200-μm intervals through the substantia nigra were stained for tyrosine hydroxylase immunoreactivity (anti-TH MAB318, 1:1,000; Chemicon) to visualize dopaminergic neurons (Fig. S5).

SI Supporting Data and Discussion

Value Coding and Uncertainty of Reward and Decision. Are there any signs that the dopamine responses possibly encoded something other than the long-term value? Dopamine neuronal activities may represent such previously reported factors as an uncertainty of reward availability (9) or decision confidence (10). Reward uncertainty reaches its peak when decision confidence is at its lowest, at a 50% probability. In our multistep choice paradigm in both monkeys, the maximum dopamine responses occurred at approximately 80% reward probability (N3 trials) and were significantly larger than the responses at approximately 50% probability (N2 trials; $P < 0.01$ in both monkeys, Mann–Whitney *U* test, Fig. 3 *B* and *C*).

To examine how the expectation of multiple future rewards and reward uncertainty are involved in our multistep actions, we conducted two kinds of control experiments. First, we manipulated uncertainty in the N3 trials of the multistep choice task, in which the maximum dopamine responses to the task start cue occurred. Only unchosen targets in the previous one or two trials appeared in the next trials, so the monkey chose one of a progressively reduced number of illuminated targets (Fig. S4*A*). Thus, just a single target appeared in the N3 trials. The reward probability improved considerably from 74% in the original task (with both chosen and unchosen targets remaining lit) to 99% in the control task. The uncertainty of reward greatly decreased from 0.83 bits in the original task to 0.09 bits in the control task (Fig. S4*B*). The dopamine responses to the task start cue in the control task were largest in the N3 trials for the first reward, were significantly reduced in the R1 trials for the second reward (Mann –Whitney *U* test, $P < 0.05$), and were very similar to the responses recorded in the original task from the same neurons. The responses in the N3 trial were not different between the two tasks (Mann–Whitney *U* test, $P = 0.89$). Therefore, reward uncertainty was not critical for the inverted V-shaped distribution of dopamine responses in the multistep choice task. A plot of the dopamine responses against trial type and reward probability was also very similar to the duration of anticipatory licking in the control and original tasks ($R = 0.95$, $P = 0.07$ in the control task; and $R = 1.00$, $P < 0.001$ in the original task, Fig. S4*B*), which indicated very similar reward expectations during the two tasks.

For the second control experiment, we asked monkeys to expect a reward in the current trial but no future rewards by using a classical conditioning paradigm in which patterns of illuminated LEDs were presented as CSs with variable reward probabilities (0%, 20%, 50%, 80%, and 100%; Fig. S5*A*). Different types of CSs were presented in a random order, so the monkeys could not predict the type of CS before its appearance. During this task, we examined the responses of another population of 47 dopamine neurons in monkeys BT and CC (Table S1). In both monkeys, dopamine responses to the CS were modulated monotonically with the increase in reward probability; a maximum decrease in firing after the CSs that associated with the lowest reward probability (P0 in monkey BT, P20 in monkey CC; Tukey–Kramer test, $P < 0.01$ in both monkeys), and a maximum increase in firing after the CS that associated with the highest reward probability (P100; $P < 0.01$ in both monkeys, except between P80 and P100; Fig. S5 *C* and *D*). The monkeys engaged in anticipatory licking after the CS presentation for gradually longer periods in parallel with an increase in reward probability (Tukey–Kramer test, $P < 0.01$ between P100 and other all trial types in monkeys BT and CC, except between P80 and P100 in monkey CC; Fig. S5 *B* and *D*), which indicated that reward expectation was modulated monotonically with the reward probability of the currently presented CS. The magnitude of the dopamine responses was significantly correlated with licking durations ($R = 0.94$, $P < 0.05$ in monkey BT; $R = 0.99$, $P < 0.05$ in monkey CC), which is consistent with previous observations (9, 11). Therefore, dopamine neurons encode the value of the present stimuli when they are not associated with multiple future rewards. On the contrary, a previous study (9) showed, under classical conditioning, that activity gradually increases until the potential time of reward, the occurrence of which is correlated with uncertainty and reaches maximum at 50% reward probability. However, we did not observe this type of activity, probably because of differences in behavioral conditioning (Fig. S5*E*). Based on the two control experiments, we can conclude that the dopamine responses during the multistep choice paradigm represented the long-term value of a series of actions as a sum of expected multiple future rewards at different times in which distant rewards are discounted.

An alternative to the long-term value coding is that dopamine neurons may respond strongly to a reliable predictor of the first of two or three rewards that are expected during the course of multistep choices. However, the monkeys still chose no-reward options in a considerable proportion of these trials (25% in monkey CC and 10% in monkey SK during the three-step choices and 21% in monkey SK and 26% in monkey BT during the two-step choices). Furthermore, the dopamine responses with inverted V-shape distributions against trial types were accurately approximated by the magnitudes of anticipatory licking, which reflected the expectation of the sum of immediate and future rewards (Figs. 2*C* and 4*B* and Fig. S4*B*). Therefore, it is unlikely that the N3 responses with the largest magnitudes reflected a special salience of the N3 cue as the first predictor of reward.

The dopamine neurons showed responses to the positive reinforcers in the R1 and R2 trials during which reward probabilities were almost 100% and thus errors were expected to be very low (Fig. S3). However, the actual probabilities of obtaining a reward were not perfect, but 3% to 6% of trials resulted in no reward (Fig. 1*C*). This appeared to be the primary reason of the small activation. Modest responses to a highly expected reward (100% probability) have also been observed previously in an instrumental conditioning task (12) and a classical conditioning task (9).

1. 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.
2. Grace AA, Bunney BS (1983) Intracellular and extracellular electrophysiology of nigral dopaminergic neurons—1. Identification and characterization. *Neuroscience* 10:301–315.
3. Schultz W (1986) Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *J Neurophysiol* 56:1439–1461.
4. Kimura M (1986) The role of primate putamen neurons in the association of sensory stimuli with movement. *Neurosci Res* 3:436–443.
5. Busemeyer JR, Diederich A (2009) *Cognitive Modeling* (Sage, Thousand Oaks, CA).
6. Lewandowsky S, Farrell S (2011) *Computational Modeling in Cognition: Principles and Practice* (Sage Publications, Inc, Thousand Oaks).
7. Daw DN (2011) Trial-by-Trial Data Analysis Using Computational Models. *Decision Making, Affect, and Learning, Attention and Performance XXIII* (Oxford Univ Press, Oxford, UK), pp 3–38.
8. Shannon CE, Weaver W (1949) *The Mathematical Theory of Communication* (Univ Illinois Press, Urbana, IL).

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

10. Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455:227–231.

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

12. 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.

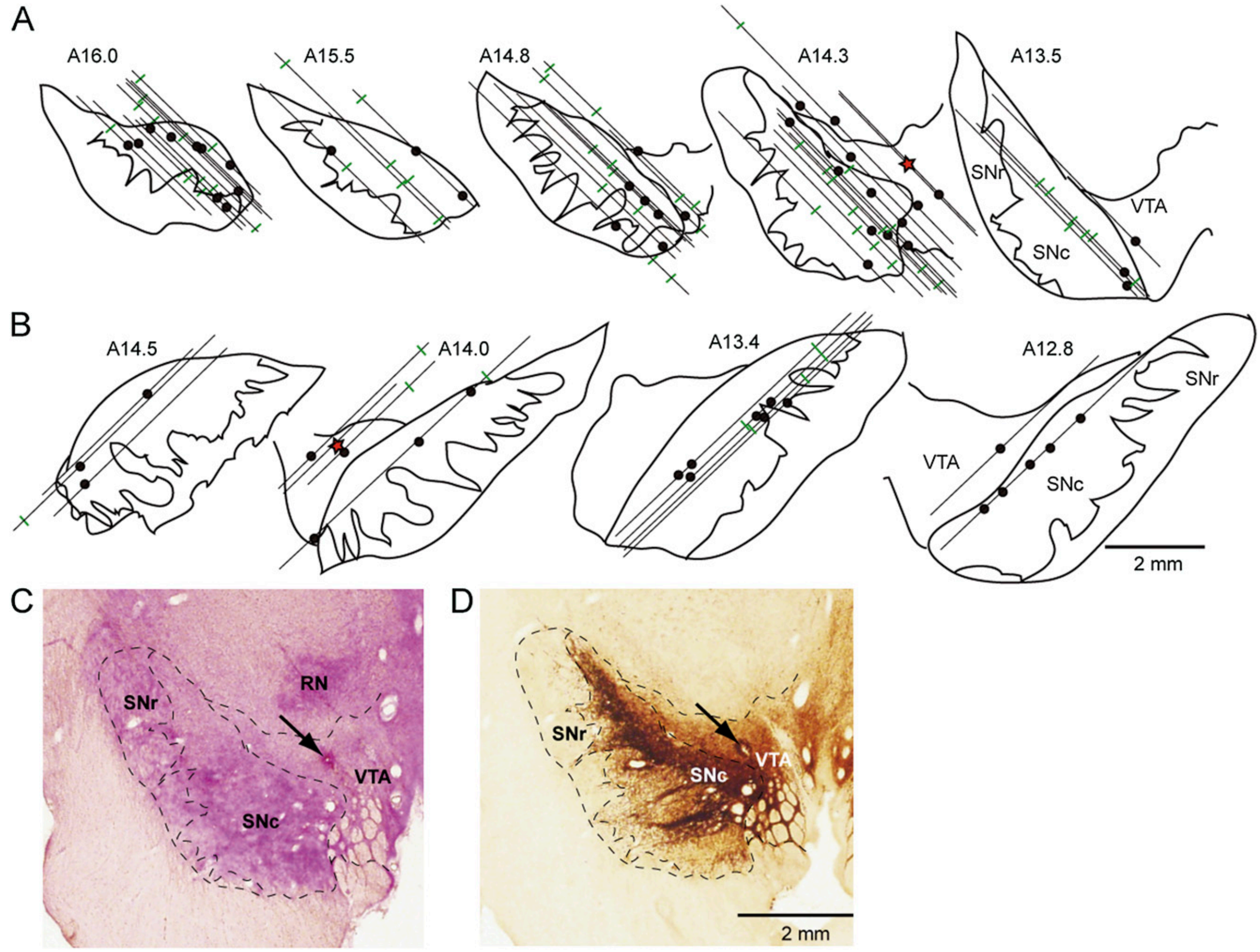

Fig. S1. Reconstruction of recording sites of dopamine neurons in the midbrain. (*A*) Histological reconstruction of the recording sites of dopamine neurons (filled circles) and nondopamine neurons (green lines) along electrode tracks in the SNc, substantia nigra pars reticulata (SNr), and VTA in the left hemisphere of monkey BT. Red stars indicate locations of electrolytic lesion marks. Numbers above each drawing indicate their distances (in mm) from interaural line. (*B*) Same as *A* but in the right hemisphere. (*C*) Nissl-stained section at the level of the SNc and substantia nigra pars reticulata and red nucleus (RN). Arrow indicates an electrolytic lesion made after recording dopamine neuron activity. (*D*) A neighboring section to *C* stained with tyrosine hydroxylase-like immunoreactivity (arrow, electrolytic lesion).

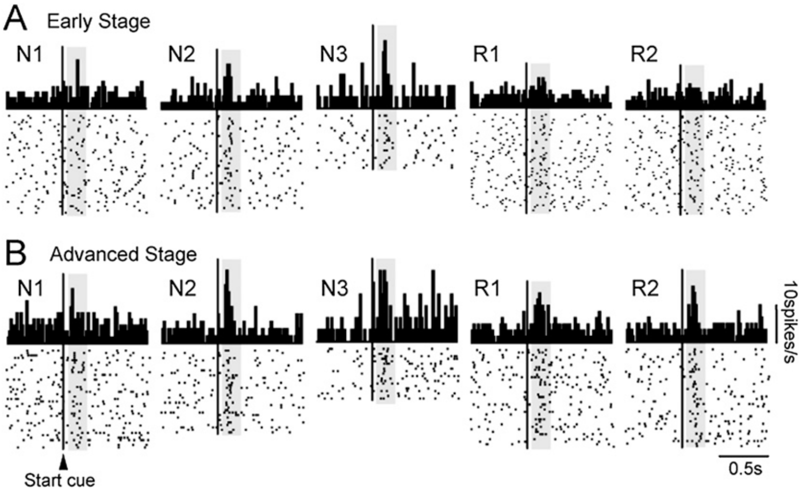

Fig. S2. Development of dopamine responses to the start cue. (*A*) Example responses of a dopamine neuron to the illumination of the start cues in individual trials on day 12 (i.e., early stage) of learning of the three-step choice task in monkey SK. The bin size of the spike density histogram is 15 ms. Hatched areas are the time windows for the analyses shown in Fig. 4*C*. (*B*) Same as *A* but for day 29 (i.e., advanced stage) of learning.

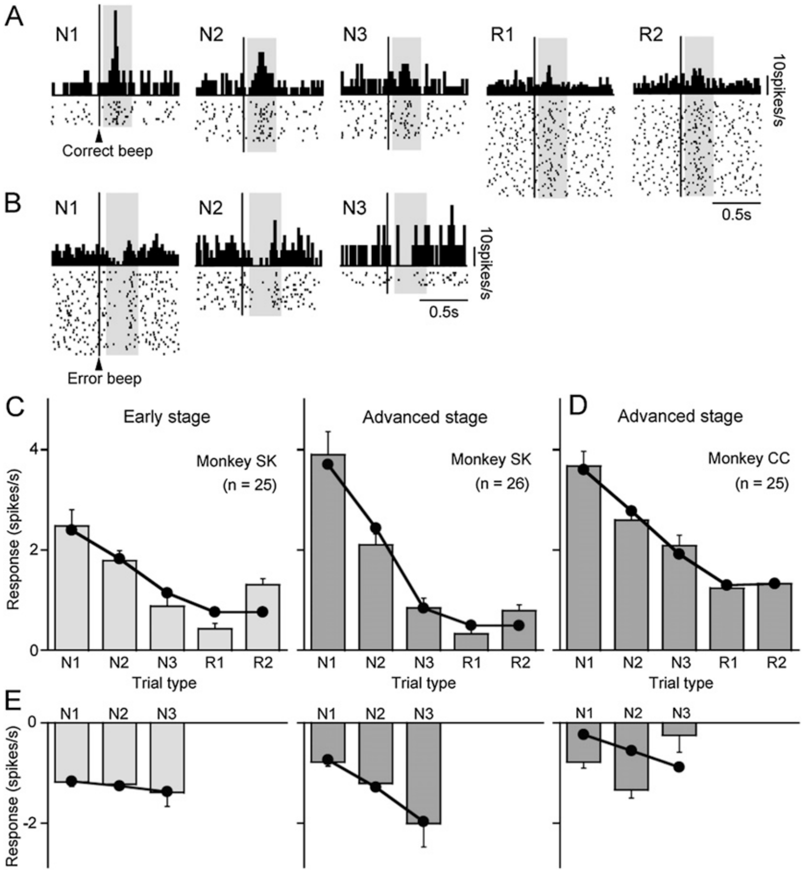

Fig. S3. Response of dopamine neurons to the positive reinforcers of the three-step choice task. (*A*) Responses of a single dopamine neuron in monkey CC. The neuronal activity is centered at the time of the beep sound after correct choices. Average discharge rates during the hatched time windows were used to measure the response amplitude (40–350 ms after the beep). The bin size of the spike density histogram is 15 ms. (*B*) Same as *A* but for responses to negative reinforcers. The neuronal activity is centered at the time of the beep sound after error choices. Time window for measurement is 80 to 410 ms after the beep. (*C*) Bar graphs of the amplitudes of neuronal responses to the positive reinforcers above the baseline activity are shown separately for trial types and learning phase (mean and SEM). Superimposed line plots are the prediction errors of long-term reward values estimated by reinforcement learning theories during the early stage (*Left*; $\gamma = 0.00$, $R = 0.90$, $P < 0.05$) and the advanced stage (*Right*; $\gamma = 0.31$, $R = 0.98$, $P < 0.01$) of learning. (*D*) Same as *B*, but for monkey CC during the advanced stage of learning ($\gamma = 0.65$, $R = 0.99$, $P < 0.01$). (*E*) Same as *C*, but for dopamine responses (decrease of discharge rate) to the negative reinforcers during the early and advanced stages.

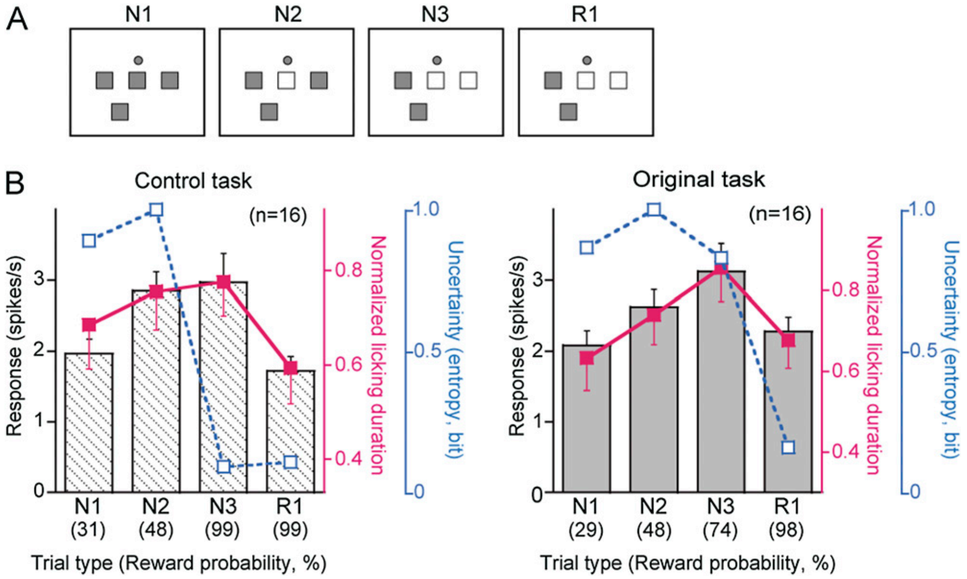

Fig. S4. Dopamine responses during the control task with reduced uncertainty for choices. (*A*) Pattern of illuminated target buttons, which progressively decreased during the exploration period. (*B*) Bar graphs of the average start cue responses above the baseline activity (mean and SEM, 70–290 ms after start cue) during the control (*Left*) and original (*Right*) tasks. The duration of normalized anticipatory licking (n = 16 sessions; 800–100 ms period before the beep sound; mean and SEM, solid red line, right axis) and the uncertainty of a reward (entropy; dashed blue line, right axis; *SI Methods*) are superimposed. Correlation of dopamine responses with anticipatory licking was R = 0.95, P = 0.07 in the control task; and R = 1.00, P < 0.001 in the original task.

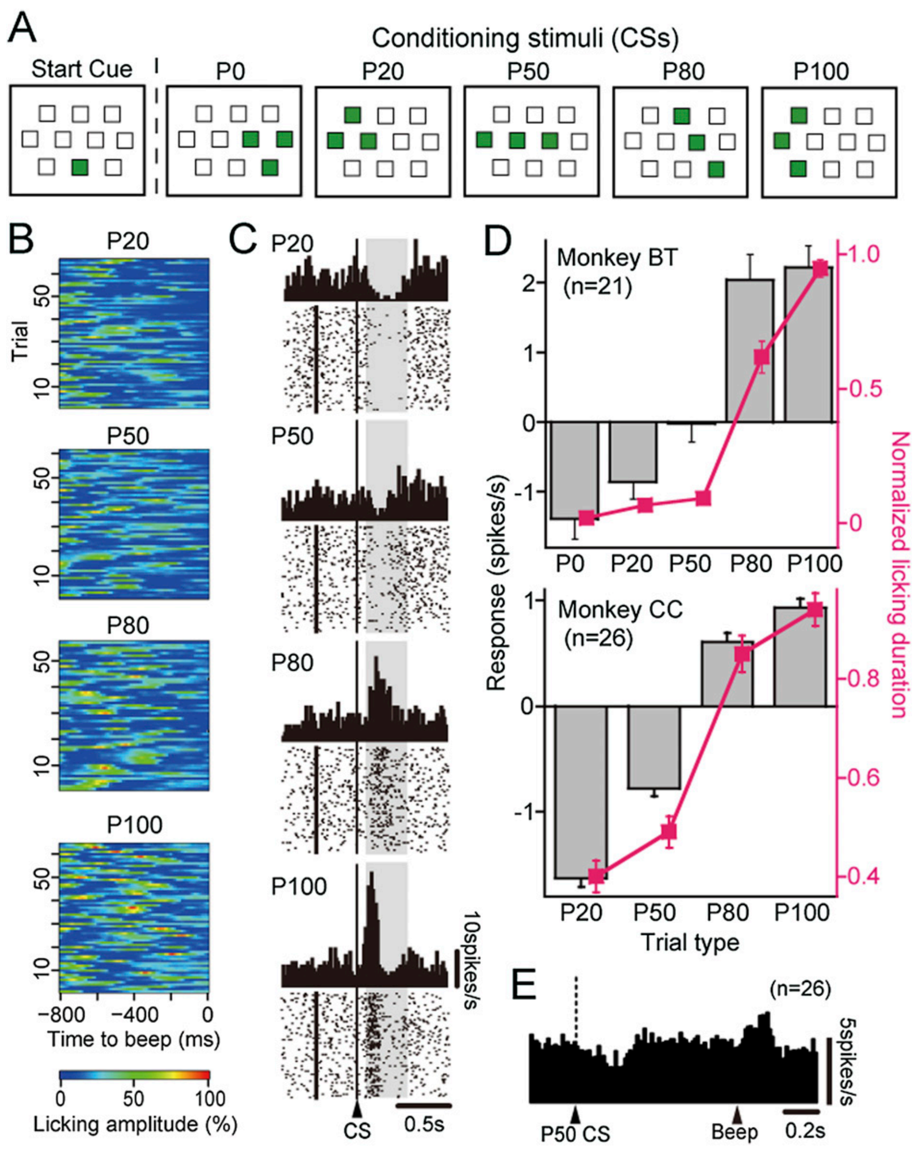

Fig. S5. Dopamine neurons encode immediate reward value of CS during the classical conditioning paradigm. (*A*) Position of the start cue LED and patterns of illuminated LEDs (three of 10 possible) in five different CSs occurring in random order for 1 s. The probabilities of reward for the different trials are shown above (P0, P20, P50, P80, and P100). (*B*) Pattern of the anticipatory licking during the recording of an example dopamine neuron illustrated in *C*. Format of illustration is the same as in Fig. 2*A*. (*C*) Responses of a dopamine neuron in monkey CC. Rasters and histograms are centered at CS onset. Vertical lines before the CS in each graph indicate the timing of the start button depression. The bin size of the spike density histograms is 15 ms. The hatched gray areas are the time windows for the analyses shown in *D*. (*D*) Bar graphs of the average neuronal responses to the CS relative to the baseline activity (mean and SEM) during time windows (80–480 ms and 40–440 ms after CS in monkey BT and monkey CC, respectively) shown in *C*, shown separately for trial types and subjects. The duration of normalized anticipatory licking (700-ms period just before the beep sound, mean and SEM, red line, right axis) is overlaid on the neural responses. Correlation of the dopamine responses with the anticipatory licking was $R = 0.94$, $P < 0.05$ in monkey BT; and $R = 0.99$, $P < 0.05$ in monkey CC. (*E*) An ensemble average histogram of the activity of 26 dopamine neurons in monkey CC. The activity is centered at the onset of the P50 CS (vertical dashed line). Bin size is 15 ms.

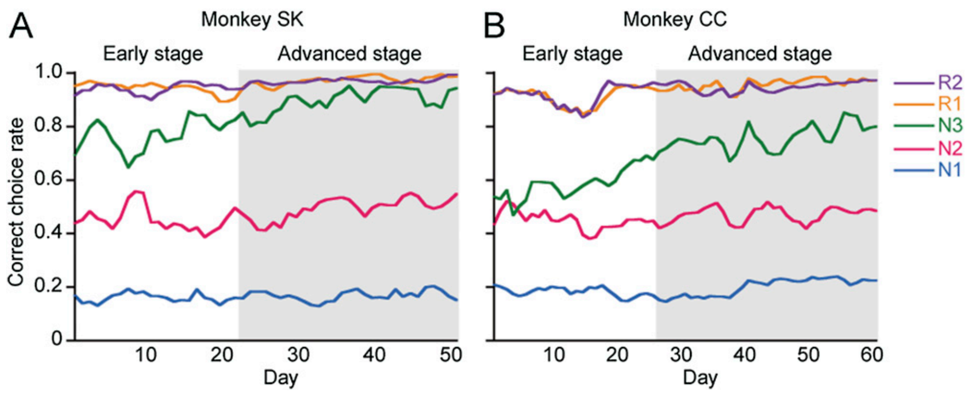

Fig. S6. Adaptive increase in correct choice rate through learning the three-step choice task. (*A*) Average correct choice rates for three consecutive days in monkey SK were plotted against the learning day. The advanced stage of learning is indicated by shading. (*B*) Same as *A* but for monkey CC.

Table S1. Database of neuronal and behavioral recordings

	Monkey			
Behavior	BT	SK	CC	Total
Two-step choice task	36	26	0	62
Three-step choice task				
Early stage	0	25	0	25
Advanced stage	0	26	25	51
Classical conditioning task	21	0	26	47
Uncertainty-reduced task	(16)	(0)	(0)	(16)
Total	57	77	51	185
Licking recording	Yes	No	Yes	—

Figures represent the number of dopamine neurons recorded in each behavioral task. Neurons recorded during the uncertainty-reduced task were included in the population recorded during the two-step choice task (the numbers in parentheses).