

Influence of Reward Delays on Responses of Dopamine Neurons

Shunsuke Kobayashi and Wolfram Schultz

Department of Physiology, Development, and Neuroscience, University of Cambridge, Cambridge CB2 3DY, United Kingdom

Psychological and microeconomic studies have shown that outcome values are discounted by imposed delays. The effect, called temporal discounting, is demonstrated typically by choice preferences for sooner smaller rewards over later larger rewards. However, it is unclear whether temporal discounting occurs during the decision process when differently delayed reward outcomes are compared or during predictions of reward delays by pavlovian conditioned stimuli without choice. To address this issue, we investigated the temporal discounting behavior in a choice situation and studied the effects of reward delay on the value signals of dopamine neurons. The choice behavior confirmed hyperbolic discounting of reward value by delays on the order of seconds. Reward delay reduced the responses of dopamine neurons to pavlovian conditioned stimuli according to a hyperbolic decay function similar to that observed in choice behavior. Moreover, the stimulus responses increased with larger reward magnitudes, suggesting that both delay and magnitude constituted viable components of dopamine value signals. In contrast, dopamine responses to the reward itself increased with longer delays, possibly reflecting temporal uncertainty and partial learning. These dopamine reward value signals might serve as useful inputs for brain mechanisms involved in economic choices between delayed rewards.

Key words: single-unit recording; dopamine; neuroeconomics; temporal discounting; preference reversal; impulsivity

Introduction

Together with magnitude and probability, timing is an important factor that determines the subjective value of reward. A classic example involves choice between a small reward that is available sooner and a larger reward that is available in the more distant future. Rats (Richards et al., 1997), pigeons (Ainslie, 1974; Rodriguez and Logue, 1988), and humans (Rodriguez and Logue, 1988) often prefer smaller reward in such a situation, which led to the idea that the value of reward is discounted by time.

Economists and psychologists have typically used two different approaches to characterize the nature of temporal discounting of reward. A standard economics model assumes that the value of future reward is discounted because of the risk involved in waiting for it (Samuelson, 1937). Subjective value of a future reward was typically formulated with exponential decay functions under assumption of a constant hazard rate corresponding to constant discounting of reward per unit time.

In contrast, behavioral psychologists found that animal choice can be well described by hyperbola-like functions. An essential property of hyperbolic discounting is that the rate of discounting is not constant over time; discounting is larger in the near than far future.

Despite intensive behavioral research, neural correlates of

temporal discounting were largely unknown until recent studies shed light on several brain structures possibly involved in the process. The human striatum and orbitofrontal cortex (OFC) showed greater hemodynamic response to immediate rather than delayed reward (McClure et al., 2004; Tanaka et al., 2004). The preference of rats for small immediate reward over larger delayed reward increases with lesions of the ventral striatum and basolateral amygdala (Cardinal et al., 2001; Winstanley et al., 2004), and decreases with excitotoxic and dopaminergic lesions of the OFC (Kheramin et al., 2004; Winstanley et al., 2004).

Midbrain dopamine neurons play a pivotal role in reward information processing. Some computational models assume that dopamine neurons incorporate the discounted sum of future rewards in their prediction error signals (Montague et al., 1996). However, there is little physiological evidence available to support the assumption of temporal discounting. Recently, Roesch et al. (2007) studied the responses of rodent dopamine neurons during an intertemporal choice task. They found that the initial phasic response of dopamine neurons reflects the more valuable option (reward of shorter delay or larger magnitude) of the available choices and the activity after decision reflects the value of the chosen option. However, it is still unclear whether temporal discounting occurs during the decision process or the decision is made by receiving delay-discounted value signals as inputs. To address this issue, we used a pavlovian conditioning task and investigated whether and how the value signals in dopamine neurons are discounted by reward delay in the absence of choice. In this way the results were comparable with previous studies that examined the effects of magnitude and probability of reward (Fiorillo et al., 2003; Tobler et al., 2005). We used an intertempo-

Received Oct. 25, 2007; revised June 14, 2008; accepted June 23, 2008.

This work was supported by the Wellcome Trust and the Cambridge Medical Research Council–Wellcome Behavioural and Clinical Neuroscience Institute. We thank P. N. Tobler, Y. P. Mikheenko, and C. Harris for discussions.

Correspondence should be addressed to Shunsuke Kobayashi, Department of Physiology, Development, and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK. E-mail: skoba-tky@umin.ac.jp.

DOI:10.1523/JNEUROSCI.1600-08.2008

Copyright © 2008 Society for Neuroscience 0270-6474/08/287837-10\$15.00/0

ral choice task to investigate the animals' behavioral valuation of reward delivered with a delay.

Materials and Methods

Subjects and surgery

We used two adult male rhesus monkeys (*Macaca mulatta*), weighing 8–9 kg. Before the recording experiments started, we implanted under general anesthesia a head holder and a chamber for unit recording. All experimental protocols were approved by the Home Office of the United Kingdom.

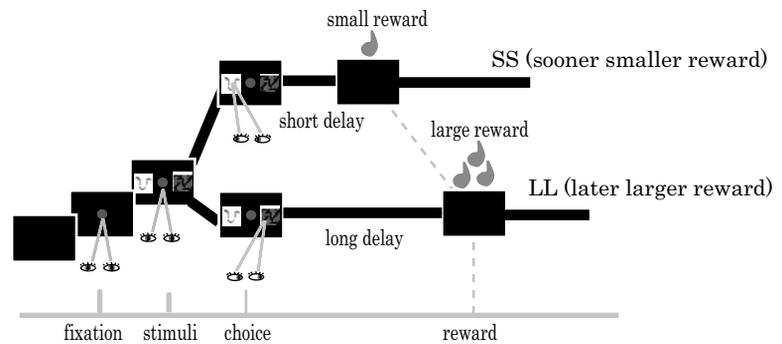
Behavioral paradigm

Pavlovian conditioning task (Fig. 1B). We presented visual stimuli on a computer display placed at 45 cm in front of the animals. Stimuli were associated with different delays (2.0, 4.0, 8.0, and 16.0 s) and magnitudes (animal A, 0.14 and 0.56 ml; animal B, 0.58 ml) of reward (a drop of water). Different complex visual stimuli were used to predict different delays and magnitudes of reward. Visual stimuli were counter-balanced between the two animals for both delay and magnitude of reward. Intertrial interval (ITI; from reward offset until next stimulus onset) was adjusted for both animals such that cycle time (reward delay from stimulus onset + ITI) was fixed at 22.0 ± 0.5 s in every trial regardless of the variable delay between a conditioned stimulus and reward. When the reward delay was 2.0 s, for example, ITI was 20.0 ± 0.5 s.

Pavlovian conditioning was started after initial training to habituate animals to sit relaxed in a primate chair inside an experiment room. In the first 3 weeks of pavlovian conditioning training, we aimed to familiarize the animals to watch the computer monitor and drink from the spout. Visual stimuli of different reward conditions were gradually introduced as training advanced during this period. From the fourth week, we trained all reward conditions randomized. Daily session was scheduled for 600 trials, but it was stopped earlier if animal started to lose motivation, for example closing eyes. In total, animal A was trained for 19,507 trials and animal B was trained for 19,543 trials in the pavlovian conditioning task before the dopamine recording was started.

Intertemporal choice task (Fig. 1A). To assess the animals' preference for different delays and magnitudes of reward, we designed the choice task, in which animals choose between sooner smaller (SS) and later larger (LL) reward. A trial of the intertemporal choice task started with the onset of the central fixation spot (1.3° in visual angle). After the animals gazed at the fixation spot for 500 ± 200 ms, two target pictures (3.6°) were presented simultaneously on both sides of the fixation spot (8.9° from the center). One target predicted SS reward and the other target predicted LL reward. The animals were required to make a choice by saccade response within 800 ms after onset of the targets. When the saccade reached a target, a red round spot appeared for 500 ms superimposed on the chosen target. Two targets remained visible on the computer monitor after the choice until the delay time associated with the chosen target elapsed and reward was delivered. The positions of two targets were randomized in every trial. The animal was not required to fixate its gaze during the reward delay. Trials were aborted on premature fixation breaks and inaccurate saccades, followed by repetition of the same trials. The ITI was adjusted such that cycle time was constant at 22.0 ± 0.5 s in all trials. When the animal chose the 2 s delayed reward, for example, ITI was 20.0 ± 0.5 s.

A Intertemporal choice task



B Pavlovian task with variable delay

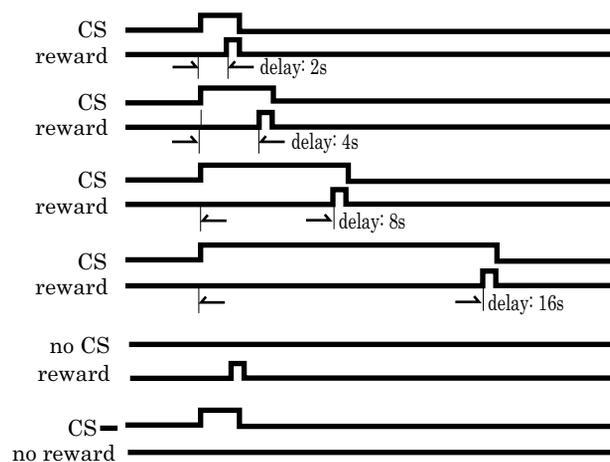


Figure 1. Experimental design. **A**, Intertemporal choice task used for behavioral testing. The animal chooses between an SS reward and an LL reward. The delay of SS and magnitude of LL were fixed. The delay of LL and magnitude of SS varied across blocks of trials. **B**, Pavlovian task used for dopamine recording. The delay of reward, which varied in every trial, was predicted by a conditioned stimulus (CS).

The visual stimuli trained in the pavlovian task as reward predictors were used as choice targets. A pair of visual stimuli served as SS and LL targets for choice, and an identical pair was tested repeatedly with left–right positions randomized within a block of 20 successful trials. Therefore, the conditions of magnitude and delay for SS and LL rewards were constant within a block. Four different magnitudes of SS reward were tested in different blocks (animal A, 0.14, 0.24, 0.35, and 0.45 ml; animal B, 0.22, 0.32, 0.41, and 0.54 ml). The SS magnitude changed increasingly or decreasingly across successive blocks, and the direction of change alternated between days. The SS delay was constant throughout the whole experiment (animal A, 2.0 s; animal B, zero delay). The LL delay changed across blocks (animal A, 4.0, 8.0, and 16.0 s; animal B, 2.0, 4.0, 8.0, and 16.0 s), whereas the LL magnitude was kept constant (animal A, 0.56 ml; animal B, 0.58 ml).

A set of 12 and 16 different blocks contained all possible combinations of SS and LL conditions for animals A and B, respectively (animal A, 4 different SS magnitudes \times 3 different LL delays; animal B, 4 different SS magnitudes \times 4 different LL delays). Animal A was tested 9 times for the complete set of 12 blocks, and animal B was tested 14 times for the complete set of 16 blocks. For both animals, two sets were tested before the neurophysiological recording and after the initial training in the pavlovian conditioning task. The rest (animal A, 7 sets; animal B, 12 sets) were interspersed with pavlovian sessions during the recording period. The data obtained were expressed as the probability of choosing the SS target, which depended on the specific magnitude and delay conditions in each of the different choice blocks.

Preference reversal test. In addition to the intertemporal choice task described above, we tested whether the animal changes preference between SS and LL over a different range of delays (preference reversal) with one animal (animal A). We changed the reward delay keeping the reward magnitude constant at 0.28 ml (SS) and 0.56 ml (LL). Three pairs of reward delay, with the constant difference of 4.0 s between short and long delays, were tested, namely 1.0 s versus 5.0 s, 2.0 s versus 6.0 s, and 6.0 s versus 10.0 s (SS vs LL). Task schedule was the same as in the intertemporal choice task described above. Each pair was tested in 10 blocks of trials, with each block consisting of 20 successful trials.

The intertemporal choice task was used to evaluate behavioral preference and construct a psychometric function of temporal discounting. We used a pavlovian conditioning task to measure response of dopamine neurons. We did not use the intertemporal choice task for this purpose because simultaneous presentation of two stimuli makes it difficult to interpret whether dopamine response reflects the value of SS, LL, or their combination. Thus, we tested dopamine response in a simple pavlovian situation and measured the effect of reward delay.

Recording procedures

Using conventional techniques of extracellular recordings *in vivo*, we studied the activity of single dopamine neurons with custom-made, movable, glass-insulated, platinum-plated tungsten microelectrodes positioned inside a metal guide cannula. Discharges from neuronal perikarya were amplified, filtered (300 Hz to 2 kHz), and converted into standard digital pulses by means of an adjustable Schmitt trigger. The following features served to attribute activity to a dopamine neuron: (1) polyphasic initially positive or negative waveforms followed by a prolonged positive component, (2) relatively long durations (1.8–3.6 ms measured at 100 Hz high-pass filter), and (3) irregular firing at low baseline frequencies (0.5–8.5 spikes/s) in sharp contrast to the high-frequency firing of neurons in the substantia nigra pars reticulata (Schultz and Romo, 1987). We also tested neuronal response to unpredicted reward (a drop of water) outside the task. Neurons that met the above three criteria typically showed a phasic activation after unexpected reward. Those that did not show reward response were excluded from the main analysis. The behavioral task was controlled by custom-made software running on a Macintosh IIfx computer (Apple). Eye position was monitored using an infrared eye tracking system at 5 ms resolution (ETL200; ISCAN). Licking was monitored with an infrared optosensor at 1 ms resolution (model V6AP; STM Sensor Technologie).

Data collection and analyses

Timings of neuronal discharges and behavioral data (eye position and licking) were stored by custom-made software running on a Macintosh IIfx computer (Apple). Off-line analysis was performed on a computer using MATLAB for Windows (MathWorks). To evaluate the effects of reward delay on choice behavior, we tested the two most widely used models of temporal discounting that assumed exponential and hyperbolic decreases of reward value by delay. The use of the term hyperbolic in this paper is meant to be qualitative, consistent with its usage in behavioral economics. There exist other, more elaborate discounting functions, such as the generalized hyperbola and summed exponential functions, which might provide better fits than single discounting functions (Corrado et al., 2005; Kable and Glimcher, 2007). However, we limited our analysis to the two simplest models that provide the best contrast to test constant versus variable discount rate over time with the same number of free parameters.

We modeled exponential discounting by the following equation:

$$V = Ae^{-kt}, \quad (1)$$

where V is the subjective value of a future reward, A is the reward magnitude, t is the delay to its receipt, and k is a constant that describes the rate of discounting. We modeled hyperbolic discounting by the following equation:

$$V = A/(1 + kt), \quad (2)$$

where V , k , and t are analogous to Equation 1 (Mazur, 1987; Richards et al., 1997). Testing of each model underwent the following two steps.

First, a testing model was formulated by fixing the discount coefficient (k in Eq. 1 or 2) at one value. The constant A in Equation 1 or 2 was derived from a constraint that the value of immediate reward was 100% (animal A, $V = 100$ at $t = 2$; animal B, $V = 100$ at $t = 0$) (Fig. 2*B, D*). Second, the current testing model provided a value of percentage discount at each delay (Fig. 2*A–D*, colored open circles), which was used to constrain the indifferent point of a psychometric curve that predicted the rate of animal's SS choice (%SS choice, ordinate) as a function of the magnitude of SS (%large reward, abscissa) (Fig. 2*A, C*). The best-fit cumulative Weibull function was obtained by the least-squares method, and goodness of fit to the choice behavior was evaluated by the coefficient of determination (R^2). By sweeping the k value in the testing model, the best model that optimizes the R^2 of behavioral data fitting was obtained. It should be noted that animal A was tested with two different magnitudes of reward, and the above procedure of model fitting was performed separately for each reward magnitude; thus, a discount coefficient (k) was obtained for each magnitude. In this way, we could compare how the rate of temporal discounting changed with reward magnitude.

To examine the relationship between dopamine activity and reward delay, we calculated Spearman's rank correlation coefficient in a window of 25 ms, which was slid across the whole trial duration in steps of 5 ms separately for each neuron. To estimate the significance of correlation, we performed a permutation test by shuffling each dataset for each time bin of 25 ms for 1000 times. A confidence interval of $p < 0.99$ was obtained based on the distribution of correlation coefficient of the shuffled dataset.

To test for exponential and hyperbolic models with dopamine activity, we fit dopamine responses to the reward-predicting stimulus to the following formulas:

$$Y = b + Ae^{-kt} \quad (3)$$

$$Y = b + A/(1 + kt), \quad (4)$$

where Y is the discharge rate, A is a constant that determines the activity at no delay (free reward), t is the length of reward delay, k is a parameter that describes discount rate, and b is a constant term to model baseline activity.

To fit the increasing reward response of dopamine neurons as a function of delay in convex shape, we chose exponential, logarithmic, and hyperbolic functions, defined as follows:

$$Y = b + A \ln(1 + kt) \quad (5)$$

$$Y = b - Ae^{-kt} \quad (6)$$

$$Y = b - A/(1 + kt), \quad (7)$$

where t and Y are variables as defined in Equations 3 and 4, k is a parameter that describes the rate of activity change by delay, and A and b are constants. The logarithmic model is based on the Weber law property of interval timing (Eq. 5). The exponential and hyperbolic models test constant and uneven rates of activity increases, respectively (Eqs. 6, 7).

The regressions were examined separately for individual neuronal activity and population-averaged activity. For both single-neuron- and population-based analyses, stimulus response was measured during 110–310 ms after stimulus onset, and reward response was measured during 80–210 ms after reward onset. The responses were normalized dividing by mean baseline activity measured 100–500 ms before stimulus onset. For the analysis of stimulus response, the response to free reward was taken as a value at zero delay ($t = 0$).

For the single-neuron-based analysis, response of each neuron in each trial was the dependent variable Y in Equations 3–7. Goodness of fit was evaluated by R^2 using the least-squares method. To examine which model gives better fit, we compared R^2 between the two models by Wilcoxon signed-rank test. For the population-based analysis, normalized activity averaged across neurons was the dependent variable. The regressions based on population-averaged activity aimed to estimate the best-fit hyperbolic function and its coefficient of discount (k) for each animal.

Histological examination

After recording was completed, animal B was killed with an overdose of pentobarbital sodium (90 mg/kg, i.v.) 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. Histological examination has not been done with animal A, because experiments are still in progress.

Results

Behavior

The two monkeys performed the intertemporal choice task (animal A, 2177 trials; animal B, 4860 trials) (Fig. 1A), in which they chose between targets that were associated with SS and LL rewards. Both animals chose SS more often when the magnitude of SS reward was larger (animal A, $p < 0.01$, $F_{(3,102)} = 100$; animal B, $p < 0.01$, $F_{(3,217)} = 171.6$) and when the delay of LL reward was longer (animal A, $p < 0.01$, $F_{(2,102)} = 80.2$; animal B, $p < 0.01$, $F_{(3,217)} = 13.8$) (Fig. 2A,C). These results indicate the animals' preference for larger magnitude and shorter delay of reward.

Indifferent choice between SS and LL implies that the two options are subjectively equivalent. For example, animal A was nearly indifferent in choosing between large (0.56 ml) 16 s delayed and small (0.14 ml) 2 s delayed rewards (Fig. 2A, the left-most red square plot). Thus, extending the delay from 2 to 16 s reduced the reward value by a factor of four. The indifferent-point measure allowed us to estimate how much reward value was discounted in each delay condition. Under the assumption of hyperbolic discounting, value was reduced to 72, 47, and 27% by 4, 8, and 16 s delay for animal A (Fig. 2A,B) and 75, 60, 42, and 27% by 2, 4, 8, and 16 s delay for animal B (Fig. 2C,D) with reference to sooner reward (2 s delayed for animal A and immediate for animal B; see Materials and Methods). We compared goodness of fit between hyperbolic and exponential models based on each set of behavioral testing (Fig. 2E). For both animals, the hyperbolic model fit better than the exponential discount model (animal A, $p < 0.05$; animal B, $p < 0.01$; Wilcoxon signed-rank test). The result confirms hyperbolic nature of temporal discounting.

Based on the better hyperbolic compared with exponential discounting, we tested preference reversal with animal A as a hallmark of hyperbolic discounting (761 trials). When a pair of stimuli indicated SS (0.28 ml delayed 1 s) and LL (0.56 ml delayed 5 s), the animal preferred SS (choice of SS, $68.9 \pm 13.0\%$, mean \pm SD). When we extended the delay of both options by 1 s without changing reward magnitude [SS (0.28 ml delayed 2 s) vs LL (0.56 ml delayed 6 s)], the animal's choice

became nearly indifferent (choice of SS, $48.6 \pm 21.9\%$). Further extension of the delay by 4 s [SS (0.28 ml delayed 6 s) vs LL (0.56 ml delayed 10 s)] reversed choice preference such that the animal chose LL more frequently than SS (choice of SS, $35.4 \pm 12.0\%$). The preferences reversed depending on reward delays, in keeping with the notion of hyperbolic discounting.

We measured animals' licking response in a pavlovian task to

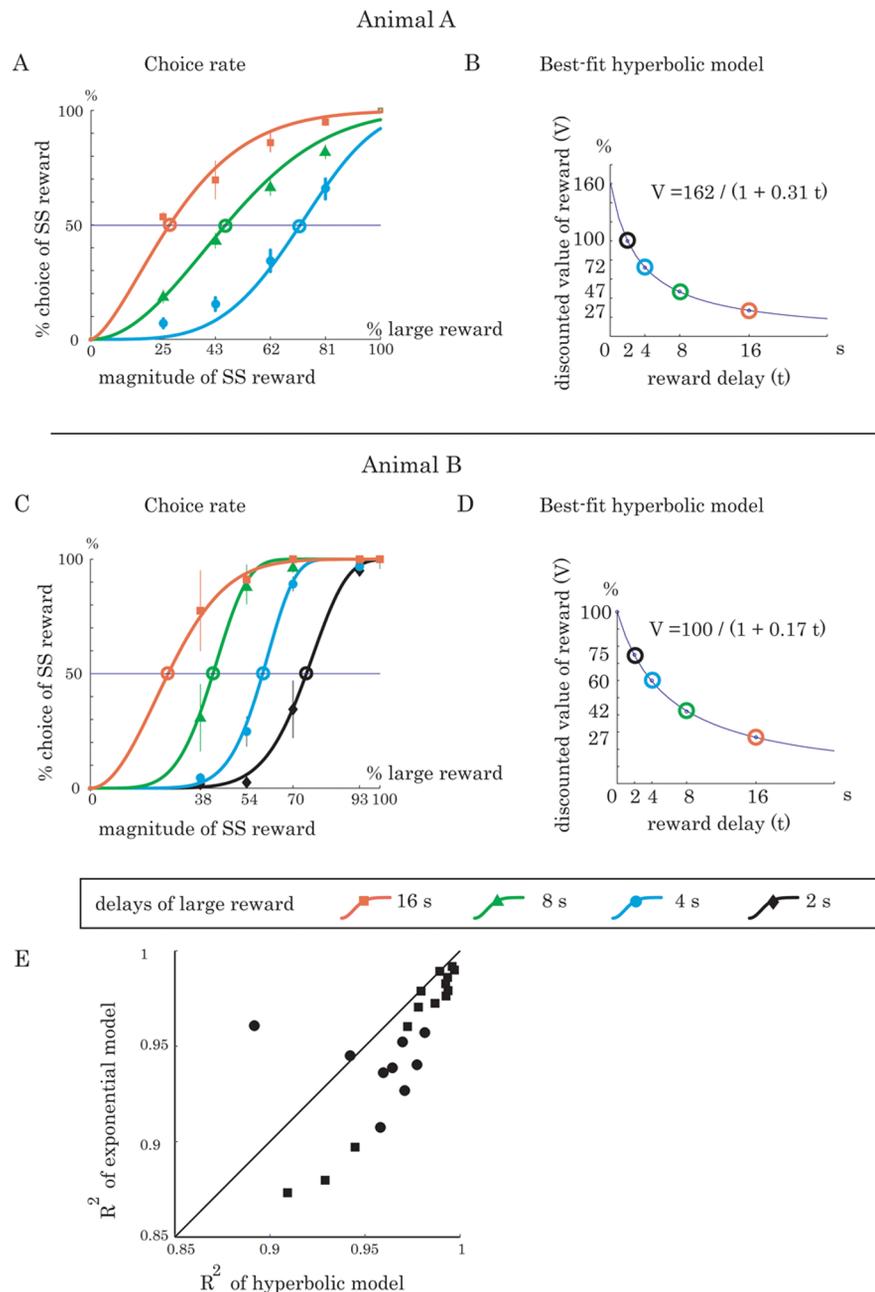


Figure 2. Impact of delay and magnitude of reward on choice behavior. **A, C**, Rate of choosing SS reward as a function of its magnitude for each animal (**A**, animal A; **C**, animal B). The magnitude of SS is plotted in percentage volume of LL reward (abscissa). The length of LL delay changed across blocks of trials (red square, 16 s; green triangle, 8 s; blue circle, 4 s; black diamond, 2 s). Curves are best-fit cumulative Weibull functions for each LL delay. Error bars represent SEM. **B, D**, Hyperbolic model that produces the least-squares error in fitting of the choice behavior (**A, C**). Value discounting (V , ordinate) is estimated relative to SS reward as a hyperbolic function of delay (t , abscissa). Because SS reward was delayed 0 s (animal A) and 2 s (animal B) from stimulus onset, the ordinate value is 100% at 0 s (**B**) and 2 s (**D**). **E**, Model fitting of behavioral choice based on individual testing sessions. Different combinations of SS and LL were tested in a set of blocks (animal A, 9 sets \times 12 different blocks; animal B, 14 sets \times 16 different blocks). Goodness of fit (R^2) of each series of datasets to hyperbolic (abscissa) and exponential (ordinate) discounting models is plotted (circles, animal A; squares, animal B; see Materials and Methods).

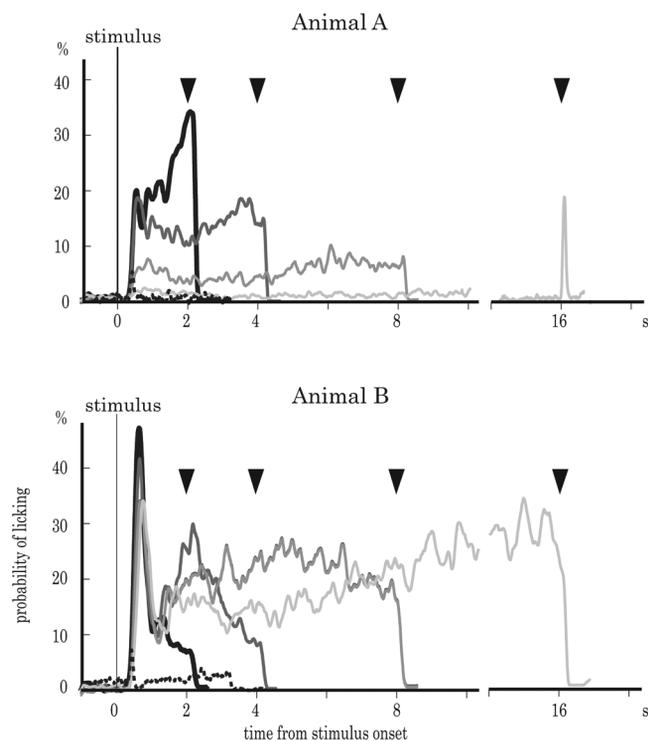


Figure 3. Probability of licking during a pavlovian task. Probability of licking of each animal is plotted as a function of time from the stimulus onset for each delay condition (2, 4, 8, and 16 s, thick black line to thinner gray lines in this order; no-reward condition, black dotted line). Triangles above indicate the onsets of reward.

the stimuli that were used in the intertemporal choice task to predict reward delays. Animals' anticipatory licking changed depending on the length of the reward delay (Fig. 3); after an initial peak immediately after stimulus presentation, the probability of licking was generally graded by the remaining time until the reward delivery. The two animals showed different patterns of licking with the 8 and 16 s delays: animal A showed little anticipatory licking with these delays, whereas animal B licked rather continuously until the time of reward delivery. These differences may be intrinsic to the animals' licking behavior and were not related in any obvious way to differences in training or testing procedures (which were very similar in these respects; see Materials and Methods). These licking differences may not reflect major differences in reward expectation, because the behaviorally expressed preferences for 8 and 16 s delayed rewards were similar for the two animals in the intertemporal choice task (Fig. 2). The probability of licking in the no-reward condition was close to zero after the initial peak (Fig. 3, dotted black line). The licking behavior of the animals may reflect the time courses of their reward expectation during the delays and the different levels of pavlovian association in each condition.

Neurophysiology

Neuronal database

We recorded single-unit activity from 107 dopamine neurons (animal A, 63 neurons; animal B, 44 neurons) during the pavlovian conditioning paradigm with variable reward delay (Fig. 1B). Baseline activity was 3.48 ± 1.78 spikes/s. Of these neurons, 88.8% (animal A, 61 neurons; animal B, 34 neurons) showed an activation response to primary reward. Eighty-seven neurons (81.3%; animal A, 54 neurons; animal B, 33 neurons) showed

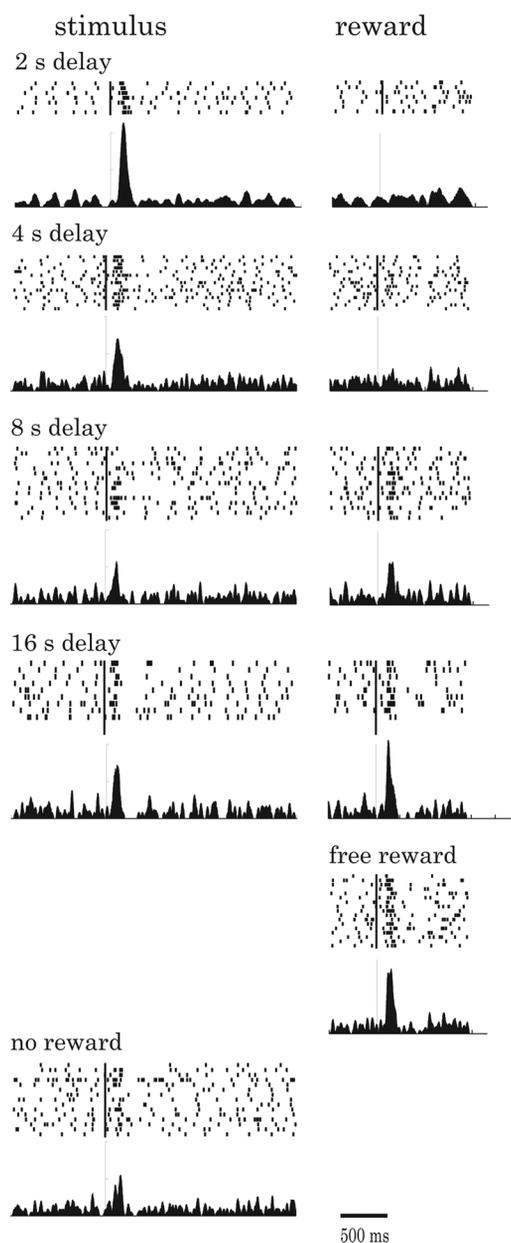


Figure 4. Example dopamine activity during a pavlovian paradigm with variable delay. Activity from a single dopamine neuron recorded from animal A is aligned to stimulus (left) and reward (right) onsets for each delay condition. For each raster plot, the sequence of trials runs from top to bottom. Black tick marks show times of neuronal impulses. Histograms show mean discharge rate in each condition. Stimulus response was generally smaller for instruction of longer delay of reward (delay conditions of 2, 4, 8, and 16 s displayed in the top 4 panels in this order). The panel labeled "free reward" is from the condition in which reward was given without prediction; hence, only reward response is displayed. The panel labeled "no reward" is from the condition in which a stimulus predicted no reward; hence, only stimulus response is displayed.

excitation to the conditioned stimuli significantly above the baseline activity level ($p < 0.05$; Wilcoxon signed-rank test).

Sensitivity of dopamine neurons to reward delay

The activity of a single dopamine neuron is illustrated in Figure 4. The magnitude of the phasic response to pavlovian conditioned stimuli decreased with the predicted reward delay, although the same amount of reward was predicted at the end of each delay. For example, the response to the stimulus that predicted reward

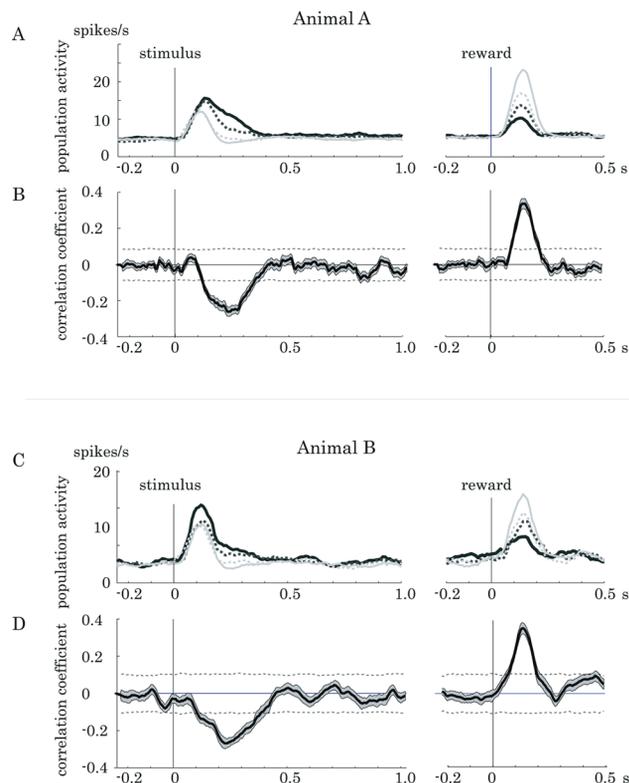


Figure 5. The effects of reward delay on population-averaged activity of dopamine neurons. **A, C**, Mean firing rate for each delay condition was averaged across the population of dopamine neurons from each animal (**A**, animal A, $n = 54$; **C**, animal B, $n = 33$), aligned to stimulus (left) and reward (right) onsets (solid black line, 2 s delay; dotted black line, 4 s delay; dotted gray line, 8 s delay; solid gray line, 16 s delay). **B, D**, Correlation coefficient between delay and dopamine activity in a sliding time window (25 ms wide window moved in 5 ms steps) was averaged across the population of dopamine neurons for each animal (**B**, animal A; **D**, animal B) as a function of time from stimulus (left) and reward (right) onsets. Shading represents SEM. Dotted lines indicate confidence interval of $p < 0.99$ based on permutation tests.

delay of 16 s was relatively small and followed by transient decrease.

Delivery of reward also activated this neuron, and the size of activation varied depending on the length of delay. Reward responses were larger after longer reward delays. The response after a reward delayed by 16 s was nearly as large as the response to unpredicted reward. Together, the responses of this dopamine neuron appeared to be influenced in opposite directions by the prediction of reward delay and the delivery of the delayed reward.

The dual influence of reward delay was also apparent in the activity averaged across 87 dopamine neurons that individually showed significant responses to both stimulus and reward. The response to the delay-predicting stimulus decreased monotonically as a function of reward delay in both animals (Fig. 5A, C, left). The changes of this response consisted of both lower initial peaks and shorter durations with longer reward delays. Conversely, the reward response increased monotonically with increasing delay with higher peaks without obvious changes in duration in both animals (Fig. 5A, C, right). We quantified the relationships between the length of delay and the magnitude of dopamine responses by calculating Spearman's rank correlation coefficient using a sliding time window. Figure 5, B and D (left), shows that the correlation coefficient of the stimulus response averaged across the 87 neurons remained insignificantly different from the chance level (horizontal dotted lines) during the initial 110–125 ms after stimulus presentation and became significantly

negative only at 125–310 ms (animal A) and 110–360 ms (animal B) (both $p < 0.01$; permutation test). Thus, the stimulus response contained an initial nondifferential component and a late differential part decreasing in amplitude with longer delays. The positive relationship of the reward response to delay was expressed by a positive correlation coefficient that exceeded chance level after 95–210 ms (animal A) and 85–180 ms (animal B) ($p < 0.01$) (Fig. 5B, D, right).

Together, these results indicate that reward delay had opposite effects on the activity of dopamine neurons: responses to reward-predicting stimuli decreased and responses to reward increased with increasing delays.

Quantitative assessment of the effects of reward delay on dopamine responses

Stimulus response. We fit the stimulus response from each dopamine neuron to exponential and hyperbolic discounting models (see Materials and Methods). Although the goodness of fit (R^2) was often similar with the two models, a hyperbolic model fit better over all ($p < 0.01$, Wilcoxon signed-rank test) (Fig. 6A). A histological examination performed on animal B showed no correlation between discount coefficient (k value) of a single dopamine neuron and its anatomical position in anterior–posterior or medial–lateral axis ($p > 0.1$, two-way ANOVA) (Fig. 7).

We examined the effect of reward magnitude together with delay in an additional 20 neurons of animal A, using small (0.14 ml) and large (0.56 ml) rewards. Normalized population activity confirmed the tendency of hyperbolic discounting, with rapid decrease in the short range of delay up to 4 s and almost no decay after 8 s for both sizes of reward (Fig. 6B; small reward, gray diamonds; large reward, black squares). The best-fitting hyperbolic model provided an estimate of activity decrease as a continuous function of reward delay (Fig. 6B, C; solid and dotted lines, hyperbolic discount curve; shading, confidence interval $p < 0.95$). The rate of discounting was larger for small reward (animal A, $k = 0.71$, $R^2 = 0.982$) than for large reward (animal A, $k = 0.34$, $R^2 = 0.986$; animal B: $k = 0.2$, $R^2 = 0.972$). The effects of magnitude and delay on the stimulus response of dopamine neurons were indistinguishable. For example, Figure 6B shows that the prediction of large reward (0.56 ml) delayed by 16 s activated dopamine neurons nearly as much as small reward (0.14 ml) delayed by 2 s. Interestingly, the animal showed similar behavioral preferences to these two reward conditions in the choice task (Fig. 2A, red line at choice indifference point). In sum, the stimulus response of dopamine neurons decreased hyperbolically with both small and large rewards, but the rate of decrease, governed by the k value, depended on reward magnitude.

Reward response. To quantify the increase of reward response with reward delay, we fit the responses to logarithmic, hyperbolic, and exponential functions (see Materials and Methods). The model fits of responses from single dopamine neurons were generally better with the hyperbolic function compared with the exponential (Fig. 6D) ($p < 0.001$) or logarithmic functions ($p < 0.001$, Wilcoxon signed-rank test). These data indicate a steeper response slope (Δ response/unit time) at shorter compared with longer delays. Figure 6, E and F, shows population activity and the best-fitting hyperbolic model [animal A, hyperbolic: $R^2 = 0.992$ (Fig. 6E); animal B, $R^2 = 0.972$ (Fig. 6F)]. The rate of activity increase based on the hyperbolic model depended on the magnitude of reward (large reward, $k = 0.1$; small reward, $k = 0.2$). These results indicate that the increases of reward response with longer reward delays conformed best to the hyperbolic model.

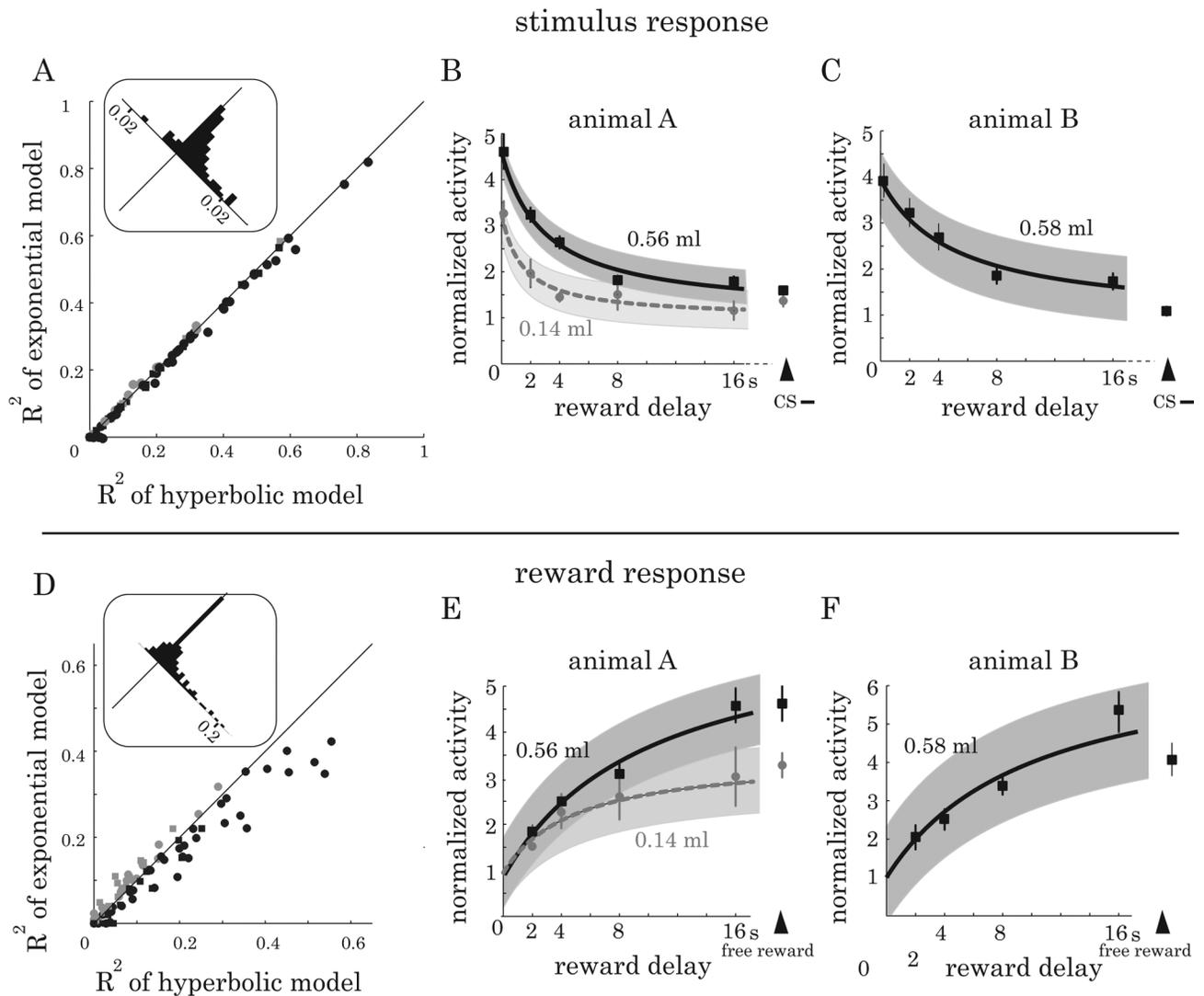


Figure 6. Hyperbolic effect of delay on dopamine activity. **A**, Goodness of fit (R^2) of stimulus response of dopamine neurons to hyperbolic (abscissa) and exponential (ordinate) models. Each symbol corresponds to data from a single neuron (black, activity that fits better to the hyperbolic model; gray, activity that fits better to the exponential model) from two monkeys (circles, animal A; squares, animal B). Most activities were plotted below the unit line, as shown in the inset histogram, indicating better fit to the hyperbolic model as a whole. **B, C**, Stimulus response was normalized with reference to baseline activity and averaged across the population for each animal (**B**, animal A; **C**, animal B). Two different magnitudes of reward were tested with animal A, and large reward was tested with animal B (black squares, large reward; gray circles, small reward). Response to free reward is plotted at zero delay, and response to the stimulus associated with no reward is plotted on the right (CS—). Error bars represent SEM. The best-fit hyperbolic curve is shown for each magnitude of reward (black solid line, large reward; gray dotted line, small reward) with confidence interval of the model ($p < 0.95$; shading). **D**, R^2 of fitting reward response of dopamine neurons into hyperbolic (abscissa) and exponential (ordinate) models (black, activities that fits better to the hyperbolic model; gray, activity that fits better to the exponential model; circle, animal A; square, animal B). **E, F**, Population-averaged normalized reward response is shown for animal A (**E**) and animal B (**F**). Conventions for different reward magnitudes are the same as in **B** and **C**. Error bars represent SEM. The curves show the best-fit hyperbolic function for each magnitude of reward.

Discussion

This study shows that reward delay influences both intertemporal choice behavior and the responses of dopamine neurons. Our psychometric measures on behavioral preferences confirmed that discounting was hyperbolic as reported in previous behavioral studies. The responses of dopamine neurons to the conditioned stimuli decreased with longer delays at a rate similar to behavioral discounting. In contrast, the dopamine response to the reward itself increased with longer reward delays. These results suggest that the dopamine responses reflect the subjective reward value discounted by delay and thus may provide useful inputs to neural mechanisms involved in intertemporal choices.

Temporal discounting behavior

Our monkeys preferred sooner to later reward. As most of the previous animal studies concluded, temporal discounting of our monkeys was well described by a hyperbolic function (Fig. 2). Comparisons with other species suggest that monkeys discount less steeply than pigeons, as steeply as rats, and more steeply than humans (Rodriguez and Logue, 1988; Myerson and Green, 1995; Richards et al., 1997; Mazur et al., 2000).

The present study demonstrated preference reversal in the intertemporal choice task, indicating that animal's preference for delayed reward is not necessarily consistent but changed depending on the range of delays (Ainslie and Herrnstein, 1981; Green and Estle, 2003). The paradoxical behavior can be explained by

uneven rate of discounting at different ranges of delay, e.g., in the form of a hyperbolic function, and/or different rate of discounting at different magnitudes of reward (Myerson and Green, 1995).

Dopamine responses to conditioned stimuli

Previous studies showed that dopamine neurons change their response to conditioned stimuli proportional to magnitude and probability of the associated reward (Fiorillo et al., 2003; Tobler et al., 2005). The present study tested the effects of reward delay as another dimension that determines the value of reward and found that dopamine responses to reward-predicting stimuli tracked the monotonic decrease of reward value with longer delays (Figs. 4–6).

Interestingly, delay discounting of the stimulus response emerged only after an initial response component that did not discriminate between reward delays. Subsequently the stimulus response varied both in amplitude and duration, becoming less prominent with longer delays. Similar changes of stimulus responses were seen previously in blocking and conditioned inhibition studies in which a late depression followed nonreward-predicting stimuli, thus curtailing and reducing the duration of the activating response (Waelti et al., 2001; Tobler et al., 2003). Given the frequently observed generalization of dopamine responses to stimuli resembling reward predictors (Schultz and Romo, 1990; Ljungberg et al., 1991, 1992), dopamine neurons might receive separate inputs for the initial activation with poor reward discrimination and the later component that reflects reward prediction more accurately (Kakade and Dayan, 2002). Thus, a generalization mechanism might partly explain the comparable levels of activation between the 16-s-delay and no-reward conditions in the present study.

The current data show that the population response of dopamine neurons decreased more steeply for the delays in the near than far future (Fig. 6*B,C*). The uneven rates of response decrease were well described by a hyperbolic function similar to behavioral discounting. Considering that the distinction between the hyperbolic and exponential models was not always striking for single neurons (Fig. 6*A*) and the rate of discounting varied considerably across neurons (Fig. 7), it is not excluded that hyperbolic discounting of the population response was partly attributable to averaging of different exponential functions from different dopamine neurons. Nevertheless, the hyperbolic model provides at least one simple and reasonable description of the subjective valuation of delayed rewards by the population of dopamine neurons.

The effect of reward magnitude on the rate of temporal discounting is often referred to as the magnitude effect, and studies of human decision making on monetary rewards generally conclude that smaller rewards are discounted more steeply than larger rewards (Myerson and Green, 1995). We found that the stimulus response of dopamine neurons also decreased more rapidly across delays for small compared with large reward.

From an ecological viewpoint, discounting of future rewards

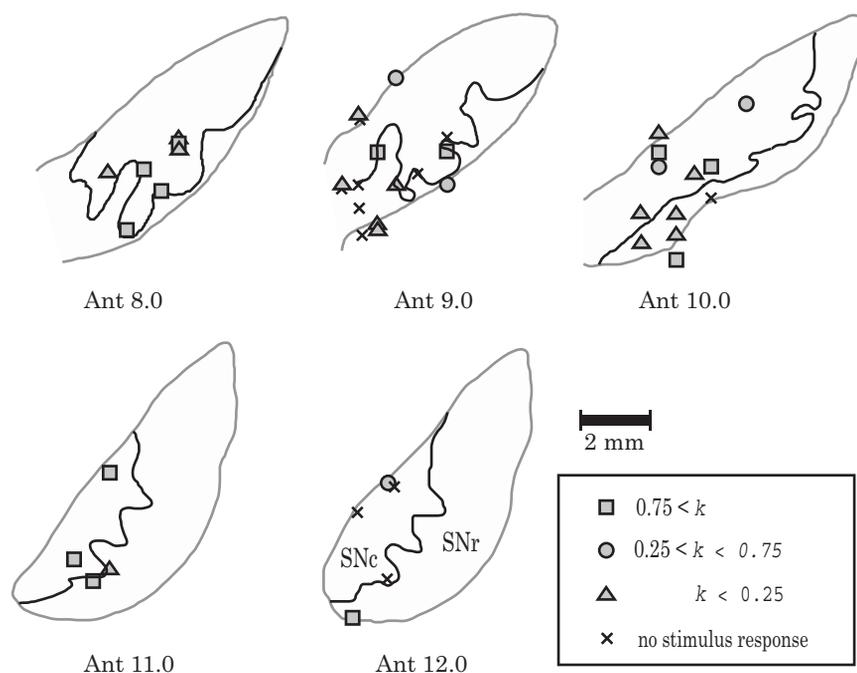


Figure 7. Histologically reconstructed positions of dopamine neurons from monkey B. Rate of discounting of stimulus response (governed by the k value in a hyperbolic function) is denoted by symbols (see inset and Materials and Methods). Neurons recorded from both hemispheres are superimposed. SNc, Substantia nigra pars compacta; SNr, substantia nigra pars reticulata; Ant 8.0–12.0, levels anterior to the interaural stereotaxic line.

may be an adaptive response to uncertainty of reward encountered in natural environment (Kagel et al., 1986). Thus, temporal discounting might share the same mechanisms as probability discounting (Green and Myerson, 2004; Hayden and Platt, 2007). Although we designed the present task without probabilistic uncertainty and with reward rate fixed by constant cycle time, further investigations are required to strictly dissociate the effects of probability and delay on dopamine activity.

How does the discounting of pavlovian value signals relate to decision making during intertemporal choice? A recent rodent study revealed that transient dopamine responses signaled the higher value of two choice stimuli regardless of the choice itself (Roesch et al., 2007). A primate single-unit study suggested different rates of discounting among dopamine-projecting areas: the striatum showed greater decay of activity by reward delay than the lateral prefrontal cortex (Kobayashi et al., 2007). Although the way these brain structures interact to make intertemporal choices is still unclear, our results suggest that temporal discounting occurs already at the pavlovian stage. It appears that dopamine neurons play a unique role of representing subjective reward value in which multiple attributes of reward, such as magnitude and delay, are integrated.

Dopamine response to reward

Contrary to its inverse effect on the stimulus response, increasing delays had an enhancing effect on the reward response in the majority of dopamine neurons (Figs. 4, 5, 6*E,F*). The dopamine response has been shown to encode a reward prediction error, i.e., the difference between the actual and predicted reward values: unexpected reward causes excitation, and omission of expected reward causes suppression of dopamine activity (Schultz et al., 1997). In the present experiment, however, the magnitude and delay of reward was fully predicted in each trial, hence in theory, no prediction error would occur on receipt of reward.

One possible explanation for our unexpected finding of larger reward responses with longer delay is temporal uncertainty; reward timing might be more difficult to predict after longer delays, hence a larger temporal prediction error would occur on receipt of reward. This hypothesis is supported by intensive research on animal timing behavior, which showed that the SD of behavioral measures varies linearly with imposed time (scalar expectancy theory) (Church and Gibbon, 1982). Our behavioral data would support the notion of weaker temporal precision in reward expectation with longer delays. Both of our animals showed wider temporal spreading of anticipatory licking while waiting for later compared with earlier rewards. However, despite temporal uncertainty, the appropriate and consistent choice preferences suggest that reward was expected overall (Fig. 2). Thus, the dopamine response appears to increase according to the larger temporal uncertainty inherent in longer delays.

Another possible explanation refers to the strength of association that might depend on the stimulus–reward interval. Animal psychology studies showed that longer stimulus–reward intervals generate weaker associations in delay conditioning (Holland, 1980; Delamater and Holland, 2008). In our study, 8–16 s of delay might be longer than the optimal interval for conditioning; thus, reward prediction might remain partial as a result of sub-optimal learning of the association. As dopamine neurons respond to the difference between the delivered reward and its prediction (Ljungberg et al., 1992; Waelti et al., 2001; Tobler et al., 2003), partial reward prediction would generate a graded positive prediction error at the time of the reward. Thus, partial reward prediction caused by weak stimulus–reward association may contribute to the currently observed reward responses after longer delays.

Computational models based on temporal difference (TD) theory reproduced the dopamine responses accurately in both temporal and associative aspects (Sutton, 1988; Houk et al., 1995; Montague et al., 1996; Schultz et al., 1997; Suri and Schultz, 1999). However, the standard TD algorithm does not accommodate differential reward responses after variable delays. Although introducing the scalar expectancy theory into a TD model is one way to explain the present data (cf. Daw et al., 2006), further experiments are required to measure the time sensitivity of dopamine neurons as a function of delay-related uncertainty. Future revisions of TD models may need to accommodate the present results on temporal delays.

Temporal discounting and impulsivity

Excessive discounting of delayed rewards leads to impulsivity, which is a key characteristic of pathological behaviors such as drug addiction, pathological gambling, and attention-deficit/hyperactivity disorder (for review, see Critchfield and Kollins, 2001). Dopamine neurotransmission has been suggested to play a role in impulsive behavior (Dalley et al., 2007). A tempting hypothesis is that temporal discounting in the dopamine system relates to behavioral impulsivity. Another popular view is that interaction between two different decision-making systems, impulsive (e.g., striatum) and self-controlled (e.g., lateral prefrontal cortex), leads to dynamic inconsistency in intertemporal choice (McClure et al., 2004; Tanaka et al., 2004). Future investigations are needed to clarify these issues, for example by comparing the rate of temporal discounting of neuronal signals between normal and impulsive subjects in the dopamine system and other reward-processing areas.

References

- Ainslie GW (1974) Impulse control in pigeons. *J Exp Anal Behav* 21:485–489.
- Ainslie GW, Herrnstein RJ (1981) Preference reversal and delayed reinforcement. *Anim Learn Behav* 9:476–482.
- Cardinal RN, Pennicott DR, Sugathapala CL, Robbins TW, Everitt BJ (2001) Impulsive choice induced in rats by lesions of the nucleus accumbens core. *Science* 292:2499–2501.
- Church RM, Gibbon J (1982) Temporal generalization. *J Exp Psychol Anim Behav Process* 8:165–186.
- Corrado GS, Sugrue LP, Seung HS, Newsome WT (2005) Linear-nonlinear-Poisson models of primate choice dynamics. *J Exp Anal Behav* 84:581–617.
- Critchfield TS, Kollins SH (2001) Temporal discounting: basic research and the analysis of socially important behavior. *J Appl Behav Anal* 34:101–122.
- Dalley JW, Fryer TD, Brichard L, Robinson ESJ, Theobald DEH, Lääne K, Peña Y, Murphy ER, Shah Y, Probst K, Abakumova I, Aigbirhio FI, Richards HK, Hong Y, Baron JC, Everitt BJ, Robbins TW (2007) Nucleus accumbens D2/3 receptors predict trait impulsivity and cocaine reinforcement. *Science* 315:1267–1270.
- Daw ND, Courville AC, Touretzky DS (2006) Representation and timing in theories of the dopamine system. *Neural Comput* 18:1637–1677.
- Delamater AR, Holland PC (2008) The influence of CS-US interval on several different indices of learning in appetitive conditioning. *J Exp Psychol Anim Behav Process* 34:202–222.
- Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299:1898–1902.
- Green L, Estle SJ (2003) Preference reversals with food and water reinforcers in rats. *J Exp Anal Behav* 79:233–242.
- Green L, Myerson J (2004) A discounting framework for choice with delayed and probabilistic rewards. *Psychol Bull* 130:769–792.
- Hayden BY, Platt ML (2007) Temporal discounting predicts risk sensitivity in rhesus macaques. *Curr Biol* 17:49–53.
- Holland PC (1980) CS-US interval as a determinant of the form of pavlovian appetitive conditioned responses. *J Exp Psychol Anim Behav Process* 6:155–174.
- 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.
- Kable JW, Glimcher PW (2007) The neural correlates of subjective value during intertemporal choice. *Nat Neurosci* 10:1625–1633.
- Kagel JH, Green L, Caraco T (1986) When foragers discount the future: constraint or adaptation? *Anim Behav* 34:271–283.
- Kakade S, Dayan P (2002) Dopamine: generalization and bonuses. *Neural Netw* 15:549–559.
- Kheramin S, Body S, Ho MY, Velázquez-Martinez DN, Bradshaw CM, Szabadi E, Deakin JFW, Anderson IM (2004) Effects of orbital prefrontal cortex dopamine depletion on inter-temporal choice: a quantitative analysis. *Psychopharmacology (Berl)* 175:206–214.
- Kobayashi S, Kawagoe R, Takikawa Y, Koizumi M, Sakagami M, Hikosaka O (2007) Functional differences between macaque prefrontal cortex and caudate nucleus during eye movements with and without reward. *Exp Brain Res* 176:341–355.
- Ljungberg T, Apicella P, Schultz W (1991) Responses of monkey midbrain dopamine neurons during delayed alternation performance. *Brain Res* 586:337–341.
- Ljungberg T, Apicella P, Schultz W (1992) Responses of monkey dopamine neurons during learning of behavioral reactions. *J Neurophysiol* 67:145–163.
- Mazur JE (1987) An adjusting procedure for studying delayed reinforcement. In: *Quantitative analyses of behavior, Vol 5* (Commons ML, Mazur JE, Nevin JA, Rachlin H, eds), Hillsdale, NJ: Erlbaum.
- Mazur JE (2000) Tradeoffs among delay, rate, and amount of reinforcement. *Behav Processes* 49:1–10.
- McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004) Separate neural systems value immediate and delayed monetary rewards. *Science* 306:503–507.
- Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16:1936–1947.

- Myerson J, Green L (1995) Discounting of delayed rewards: models of individual choice. *J Exp Anal Behav* 64:263–276.
- Richards JB, Mitchell SH, de Wit H, Seiden LS (1997) Determination of discount functions in rats with an adjusting-amount procedure. *J Exp Anal Behav* 67:353–366.
- Rodriguez ML, Logue AW (1988) Adjusting delay to reinforcement: comparing choice in pigeons and humans. *J Exp Psychol Anim Behav Process* 14:105–117.
- Roesch MR, Calu DJ, Schoenbaum G (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat Neurosci* 10:1615–1624.
- Samuelson PA (1937) Some aspects of the pure theory of capital. *Q J Economics* 51:469–496.
- 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, Romo R (1990) Dopamine neurons of the monkey midbrain: contingencies of responses to stimuli eliciting immediate behavioral reactions. *J Neurophysiol* 63:607–624.
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599.
- Suri RE, Schultz W (1999) A neural network with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience* 91:871–890.
- Sutton RS (1988) Learning to predict by the method of temporal differences. *Machine Learning* 3:9–44.
- Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, Yamawaki S (2004) Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat Neurosci* 7:887–893.
- Tobler PN, Dickinson A, Schultz W (2003) Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *J Neurosci* 23:10402–10410.
- Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. *Science* 307:1642–1645.
- Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412:43–48.
- Winstanley CA, Theobald DEH, Cardinal RN, Robbins TW (2004) Contrasting roles of basolateral amygdala and orbitofrontal cortex in impulsive choice. *J Neurosci* 24:4718–4722.