Tonic firing mode of midbrain dopamine neurons continuously tracks reward values changing moment-by-moment

Yawei Wang¹, Osamu Toyoshima¹, Jun Kunitatsu¹⁻³, Hiroshi Yamada¹⁻³, Masayuki Matsumoto^{1-3*}

¹ Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan

² Division of Biomedical Science, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan

³ Transborder Medical Research Center, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan

* Correspondence: mmatsumoto@md.tsukuba.ac.jp

Abstract

Appropriate actions are taken based on the values of future rewards. The phasic activity of midbrain dopamine neurons signals these values. Because reward values often change over time, even on a subsecond-by-subsecond basis, appropriate action selection requires continuous value monitoring. However, the phasic dopamine activity, which is sporadic and has a short duration, likely fails continuous monitoring. Here, we demonstrate a tonic firing mode of dopamine neurons that effectively tracks changing reward values. We recorded dopamine neuron activity in monkeys during a Pavlovian procedure in which the value of a cued reward gradually increased or decreased. Dopamine neurons tonically increased and decreased their activity as the reward value changed. This tonic activity was evoked more strongly by non-burst spikes than burst spikes producing a conventional phasic activity. Our findings suggest that dopamine neurons change their firing mode to effectively signal reward values, which could underlie action selection in changing environments.

Introduction

We often experience situations in which the value of future rewards changes on a moment-by-moment basis. For example, as a bartender pours a cocktail into your glass, the amount (i.e., the value) of the cocktail that you will receive gradually increases. Midbrain dopamine neurons have been considered to function as a key neural substrate that signals reward values and regulates reward-seeking behavior. These neurons respond to rewards and reward-predicting cues with phasic activity (Bayer and Glimcher, 2005; Cohen et al., 2012; Kawagoe et al., 2004; Morris et al., 2004; Satoh et al., 2003; Schultz, 1998). An influential theory proposed that such phasic activity of dopamine neurons represents reward prediction errors, discrepancies between obtained and predicted reward values, in reinforcement learning (Doya, 2002; Montague et al., 1996; Schultz et al., 1997), and the reward-evoked phasic activity of dopamine neurons has been shown to regulate this type of learning (Chang et al., 2016; Stauffer et al., 2016; Steinberg et al., 2013; Tsai et al., 2009). In addition, the cue-evoked phasic activity of dopamine neurons has been reported to reflect the value of cued rewards (Matsumoto and Hikosaka, 2009; Roesch et al., 2007; Tobler et al., 2005), and has recently been demonstrated to influence behavior associated with cued rewards (Maes et al., 2020; Morrens et al., 2020).

Although a wealth of studies have revealed roles of the phasic dopamine neuron activity in reward processing, it is unclear whether the same principle is applicable to real-life situations. Indeed, if reward values gradually change moment-by-moment as in the cocktail situation described above, the phasic activity of dopamine neurons, which is sporadic and has a short duration (around a couple of hundred milliseconds), likely fails to continuously track these changing values. Therefore, the phasic activity seems to be unable to properly regulate reward-seeking behavior in changing environments in which reward values fluctuate. To address whether and how dopamine neurons signal changing reward values, we designed a Pavlovian procedure in which a visual cue indicated the value (i.e., the amount) of a liquid reward that gradually increased or decreased moment-by-moment. We recorded single-unit activity from dopamine neurons while macaque monkeys were conditioned with this Pavlovian procedure. Here, we demonstrate a tonic firing mode of dopamine neurons that effectively tracks changing reward values.

Results

Pavlovian procedure in which the value of a cued reward gradually changed

We designed a Pavlovian procedure with three different conditions (*Figure 1A-C*) and applied it to two macaque monkeys (monkey H and monkey P). In each condition, a

bar stimulus with red and green areas was presented as a conditioned stimulus (CS) after a fixation period. The size of the green area indicated the amount (i.e., the value) of a liquid reward that the monkey would receive. The larger the green area became, the larger the associated reward value. In the first condition (value-increase condition, *Figure 1A*), the green area was minimal at the beginning and gradually increased ($3.8^\circ/\text{s}$), i.e., the reward value gradually increased (0.082 ml/s) (*Video 1*). The gradual increase randomly stopped within 2450 ms after the onset of the bar stimulus. In the second condition (value-decrease condition, *Figure 1B*), the green area was maximal at the beginning and gradually decreased ($3.8^\circ/\text{s}$), i.e., the reward value gradually decreased (0.082 ml/s) (*Video 2*). The gradual decrease randomly stopped within 2450 ms after the bar onset. In the third condition (value-fixed condition, *Figure 1C*), the size of the green area did not change and was instead fixed at the minimum, half, or maximum that predicted a 0.1-ml, 0.2-ml, or 0.3-ml reward, respectively. In addition to these CSs, a CS with a longitudinally halved green area was included, and was randomly followed by a 0.1- to 0.3-ml reward (i.e., the reward value was uncertain). In each condition, the monkeys received the reward indicated by the green area immediately after the offset of the bar stimulus. The delay between the CS onset and the reward delivery was fixed at 2850 ms in each condition and did not change across trials. Each condition consisted of

a block of 50 trials and was repeated once or more for each recording session (*Figure 1D*).

To examine whether the monkeys accurately predicted the reward values according to the size of the green area, we applied a choice task in which two bar stimuli with red and green areas were presented on the right and left sides of a fixation point (*Figure 1E*). The monkeys were required to choose one of the bar stimuli by making a saccade. These bar stimuli were identical to those used in the Pavlovian procedure except that the green area did not increase or decrease. The size of the green area indicated the reward value and was randomly assigned to each bar stimulus. We observed that the monkeys were more likely to choose the bar stimulus with the larger green area, i.e., the logistic regression slope between the choice rate of the right bar stimulus and the value difference (right bar's value – left bar's value) was significantly larger than zero (monkey H: $n = 85$ sessions, mean $\pm$ SD = 0.94 ± 1.64 , $p = 1.17 \times 10^{-15}$; monkey P: $n = 68$ sessions, mean $\pm$ SD = 0.98 ± 1.65 , $p = 7.64 \times 10^{-13}$; Wilcoxon signed-rank test) (*Figure 1F*). This suggests that the monkeys correctly predicted the reward values according to the size of the green area.

Tonic activity of dopamine neurons tracking gradually changing reward values

While the monkeys participated in the Pavlovian procedure, we recorded single-unit activity from 99 dopamine neurons (monkey H, $n = 46$; monkey P, $n = 53$) in and around the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA) (*Figure 2-figure supplement 1*, also see *Figure 2-figure supplement 2* for the characteristic spike waveforms of the recorded dopamine neurons). Here, we presumed that a possible way for dopamine neurons to signal changing reward values is to track the changing values by tonically increasing and decreasing their activity in parallel with the value change. Indeed, an example dopamine neuron tonically increased its activity as the reward value gradually increased in the value-increase condition (left in *Figure 2A*). This tonic increase in activity was clearly in contrast with the phasic responses to the CSs in the value-fixed condition, in which the reward value did not change (right in *Figure 2A*). This neuron did not exhibit clear activity modulation as the reward value gradually decreased in the value-decrease condition (middle in *Figure 2A*). Another example dopamine neuron tonically decreased its activity in the value-decrease condition (middle in *Figure 2B*), but it did not exhibit clear activity modulation in the value-increase condition (left in *Figure 2B*). This neuron also displayed phasic responses to the CSs in the value-fixed condition (right in *Figure 2B*).

To quantify tonic activity changes in the value-increase and value-decrease

conditions, we calculated the slope of the regression line between firing rate and time for each neuron (thick gray lines in the left and middle columns of *Figure 2A, B*). In the value-increase condition, 19 of the 99 dopamine neurons exhibited a significantly positive slope ($p < 0.05$; linear regression analysis), and, on average, the regression slope was significantly larger than zero ($n = 99$ neurons, mean $\pm$ SD = 0.046 ± 0.17 , $p = 0.029$; Wilcoxon signed-rank test) (*Figure 2C*). The averaged activity of the 19 dopamine neurons exhibited a clear tonic increase as the reward value gradually increased (left in *Figure 2E*), and displayed value-representing phasic responses to the CSs in the value-fixed condition (significantly positive regression coefficient between firing rate and value: $n = 19$ neurons, mean $\pm$ SD = 1.61 ± 1.42 , $p = 6.25 \times 10^{-4}$; Wilcoxon signed-rank test) (right in *Figure 2E*). In the value-decrease condition, 15 of the 99 dopamine neurons exhibited a significantly negative slope ($p < 0.05$; linear regression analysis), and, on average, the regression slope was significantly smaller than zero ($n = 99$ neurons, mean $\pm$ SD = -0.048 ± 0.18 , $p = 0.01$; Wilcoxon signed-rank test) (*Figure 2D*). The averaged activity of the 15 dopamine neurons exhibited a clear tonic decrease as the reward value gradually decreased (left in *Figure 2F*), and displayed value-representing phasic responses to the CSs in the value-fixed condition (significantly positive regression coefficient between firing rate and value: $n = 15$ neurons, mean $\pm$ SD = 2.10 ± 0.97 , $p =$

6.10 × 10⁻⁵; Wilcoxon signed-rank test) (right in *Figure 2F*). Notably, single dopamine neurons rarely exhibited both tonic increase and decrease in activity in the value-increase and value-decrease conditions, respectively (no significant correlation between the regression slopes in the value-increase and value-decrease conditions: n = 99 neurons, r = 0.048, p = 0.64; correlation analysis) (*Figure 2-figure supplement 3*). These results suggest that a subset of dopamine neurons monitored the gradually increasing and decreasing reward values by tonically changing their activity, although the neurons did not signal these values in a unified manner.

A previous study reported that dopamine neurons exhibited a tonic increase in activity when future rewards were uncertain (Fiorillo et al., 2003). Thus, the tonic increase in activity, which we observed in the value-increase condition, might be accounted for by an uncertainty-evoked activity increase, because the reward value was uncertain until the green area stopped increasing in this condition. However, in the value-fixed condition, dopamine neurons did not exhibit a tonic increase in activity when the CS did not predict the future reward with certainty (dotted curves in the right column of *Figure 2E, F*), suggesting that the tonic increase in activity in the value-increase condition was not evoked by uncertainty.

Phasic activity of dopamine neurons evoked when the reward value stopped changing

We have observed the tonic activity of dopamine neurons that gradually increased and decreased as the reward value increased and decreased, respectively. We next found unique modulations of dopamine neuron activity that were evoked when the reward value stopped increasing and decreasing. An example dopamine neuron phasically decreased its activity when the value stopped increasing in the value-increase condition (left in *Figure 3A*), and phasically increased its activity when the value stopped decreasing in the value-decrease condition (right in *Figure 3A*). Such phasic decrease and increase in activity were generally observed across the recorded dopamine neurons, and seem to be accounted for by conventional phasic dopamine inhibitory and excitatory responses to undesirable and desirable events. That is, the stopping of the value increase was undesirable and evoked the phasic decrease in dopamine neuron activity, whereas the stopping of the value decrease was desirable and evoked the phasic increase in dopamine neuron activity.

Notably, the earlier the reward value stopped increasing (i.e., the smaller the reward value), the more strongly the phasic activity decreased (significantly positive regression coefficient between firing rate and reward value: $n = 99$ neurons, $\text{mean} \pm \text{SD} = 0.31 \pm$

0.81, $p = 1.97 \times 10^{-5}$; Wilcoxon signed-rank test) (*Figure 3B*, also see left in *Figure 3D* for the averaged activity of 37 dopamine neurons with a significantly positive regression coefficient). Furthermore, the earlier the reward value stopped decreasing (i.e., the larger the reward value), the more strongly the phasic activity increased (significantly positive regression coefficient between firing rate and reward value: $n = 99$ neurons, mean $\pm$ SD = 0.59 ± 1.01 , $p = 7.66 \times 10^{-8}$; Wilcoxon signed-rank test) (*Figure 3C*, also see right in *Figure 3D* for the averaged activity of 27 dopamine neurons with a significantly positive regression coefficient). Thus, the phasic decrease and increase in dopamine neuron activity likely reflected the reward value that was fixed when the green area stopped changing.

Relationship between the tonic and phasic dopamine neuron activities

So far, we have reported that dopamine neurons exhibited not only the phasic activity but also the tonic one to effectively signal reward values. We next examined the relationship between the tonic and phasic activities. Even if dopamine neurons exhibited a cue-aligned phasic activity that represented the value of cued reward in the value-fixed condition, these neurons did not necessarily display a tonic increase in activity (i.e., a positive regression slope between firing rate and time) in the value-increase condition

(no significant correlation between the regression coefficient and the regression slope: $n = 99$ neurons, $r = -0.083$, $p = 0.42$; correlation analysis) (*Figure 4A*) or a tonic decrease in activity (i.e., a negative regression slope between firing rate and time) in the value-decrease condition (no significant correlation between the regression coefficient and the regression slope: $n = 99$ neurons, $r = -0.003$, $p = 0.97$; correlation analysis) (*Figure 4B*).

Furthermore, even if dopamine neurons exhibited a stop-aligned phasic activity decrease that reflected the reward value in the value-increase condition, these neurons did not necessarily display a tonic decrease in activity (i.e., a negative regression slope between firing rate and time) in the value-decrease condition (no significant correlation between the regression coefficient and the regression slope: $n = 99$ neurons, $r = -0.069$, $p = 0.50$; correlation analysis) (*Figure 4C*). Thus, there was no clear relationship between the phasic and tonic activity decreases. In addition, even if dopamine neurons exhibited a stop-aligned phasic activity increase that reflected the reward value in the value-decrease condition, these neurons did not necessarily display a tonic increase in activity (i.e., a positive regression slope between firing rate and time) in the value-increase (no significant correlation between the regression coefficient and the regression slope: $n = 99$ neurons, $r = 0.061$, $p = 0.55$; correlation analysis) (*Figure 4D*). Thus, there was no clear relationship between the phasic and tonic activity increases. Taken together, the

above results suggest that the tonic activity tracking the changing reward values was evoked independently of the phasic activity that represent the static, unchanged reward values.

Burst and non-burst spike firing modes of dopamine neurons

Dopamine neurons are known to exhibit two different firing modes: burst spike firing and non-burst spike firing (Grace and Bunney, 1984a, b). The phasic activity increase of these neurons, which has been shown to reflect reward prediction errors, is produced by burst spike firing (Bayer et al., 2007). Here, we attempted to determine which firing mode produced the tonic activity increase of dopamine neurons. The burst and non-burst spike firing modes were identified based on interspike intervals (ISIs) using previously established criteria (Grace and Bunney, 1984a). These criteria were 1) an ISI of 80 ms or less to identify the onset of the burst spike firing mode, and 2) an ISI of more than 160 ms to identify the end of the firing mode. We calculated the averaged activity of the 19 dopamine neurons that exhibited a tonic increase in activity in the value-increase condition separately for burst and non-burst spikes, and found that non-burst spike firing in these neurons increased tonically as the reward value gradually increased (left in *Figure 5A*). On the other hand, burst spike firing in these neurons increased when the

reward value came close to the maximum value near the end of the value-increase period. To statistically test this trend, we divided the calculation time window, which we used to detect the significantly positive regression slope of the 19 dopamine neurons, into initial, middle, and late periods (horizontal white bars in the left column of *Figure 5A*). We found that, while non-burst spike firing exhibited a significantly larger firing rate (original firing rate – baseline firing rate) compared with burst spike firing in the initial and middle periods, there was no significant difference in the late period (initial period: $p = 0.006$; middle period: $p = 0.018$; late period: $p = 0.43$; bootstrap test with 1000 repetitions). Notably, these neurons exhibited a phasic increase in activity only with burst spike firing and not with non-burst spike firing in response to the CS predicting the large reward in the value-fixed condition (right in *Figure 5A*). In addition, the phasic increase in activity evoked when the reward value stopped decreasing in the value-decrease condition was also produced only by burst spike firing (*Figure 5B*). Taken together, these data indicate that, while the conventional phasic activity of dopamine neurons signaling the static, unchanged reward values was evoked by burst spike firing, the tonic activity tracking the changing reward values was produced more largely by non-burst spike firing.

Discussion

Midbrain dopamine neurons respond to reward-predicting cues with phasic activity, and this phasic activity has been shown to represent the value of cued rewards (Matsumoto and Hikosaka, 2009; Roesch et al., 2007; Tobler et al., 2005). On the other hand, using a Pavlovian procedure in which a reward value gradually increased or decreased moment-by-moment, we found that a subset of dopamine neurons tracked the changing reward value by a tonic increase and decrease in activity rather than by the phasic activity. These neurons exhibited a conventional value-representing phasic activity when the reward value was static and when the value stopped changing. While the phasic activity was evoked by burst spike firing, the tonic activity tracking the changing reward value was produced not only by burst spike firing but also by non-burst spike firing. Particularly, the tonic activity was more largely produced by non-burst spike firing than burst spike firing.

What variable does the tonic activity of dopamine neurons represent?

Although the phasic activity of dopamine neurons has attracted attention for its crucial roles in reward processing, recent studies using freely-moving rodents reported that dopamine release in the striatum exhibited a tonic increase (or ramp) as animals approached rewards (Engelhard et al., 2019; Hamid et al., 2016; Howe et al., 2013;

262 Mohebi et al., 2019). Such tonic increase has not been observed in dopamine neuron
 263 spike activity, at least in the VTA (Mohebi et al., 2019). Although the phasic activity of
 264 dopamine neurons has long been thought to represent reward prediction errors, the tonic
 265 increase in dopamine release has been proposed to represent “state values” that
 266 gradually increase as animals spatially and temporally approach rewards (Hamid et al.,
 267 2016; Mohebi et al., 2019). However, it is not immediately clear whether the tonic
 268 increase and decrease in dopamine neuron spike activity, which we observed in head-
 269 fixed monkeys, represent state values or reward prediction errors. Because both the
 270 state values and reward prediction errors gradually changed in the same fashion in our
 271 Pavlovian procedure, we were unable to definitely identify which variable the tonic activity
 272 of dopamine neurons represented. On the other hand, we observed that, even if
 273 dopamine neurons exhibited a strong phasic activity, these neurons did not necessarily
 274 display a tonic increase or decrease in activity. That is, the tonic activity occurred
 275 independently of the phasic activity that has been thought to represent reward prediction
 276 errors. This suggests that the tonic activity might not represent reward prediction errors,
 277 and appears to be in accord with a recent perspective that dopamine neurons represent
 278 not only reward prediction errors but also multiple signals related to various events and
 279 behaviors, including punishments, movements, subjective sensory experience, response

inhibition, decision-making and working memory (Brischoux et al., 2009; da Silva et al., 2018; de Lafuente and Romo, 2011; Engelhard et al., 2019; Howard et al., 2017; Howe and Dombeck, 2016; Matsumoto and Hikosaka, 2009; Matsumoto and Takada, 2013; Menegas et al., 2018; Ogasawara et al., 2018; Yun et al., 2020).

An attribute that might cause the tonic increase and decrease in dopamine neuron activity is the motion of the green area. Since the green area dynamically increased and decreased in the value-increase and value-decrease conditions, respectively, in which we observed the tonic increase and decrease in dopamine neuron activity, it might be possible that these motions of the green area caused the tonic activity changes irrespective of the changing reward value. On the other hand, a previous study in monkeys observed that dopamine neurons exhibited a phasic activation but not tonic activity change in response to moving-dot images (Nomoto et al., 2010). According to this observation, the motion itself was unlikely to cause the tonic increase and decrease in dopamine neuron activity. Another attribute that might cause the tonic activity changes is the time distance to the reward delivery. This time distance became shorter as the green area increased and decreased. However, dopamine neurons exhibited the opposite modulations (i.e., the tonic increase and decrease in activity) in the value-increase and value-decrease conditions in which the time distance to the reward delivery

equally became shorter as the green area changed. Therefore, the time distance was also unlikely to cause the tonic increase and decrease in dopamine neuron activity.

What roles does the tonic activity of dopamine neurons play in behavior?

While the phasic dopamine activity, which is presumed to represent reward prediction errors, has been considered to provide teaching signals in reinforcement learning (Doya, 2002; Montague et al., 1996; Schultz et al., 1997), the tonic dopamine activity has been proposed to regulate motivation (Cagniard et al., 2006; Niv, 2007). A theory called “incentive salience theory” proposes that dopamine signals assign incentive values to goals or actions of behavioral processes, and motivate actions aimed at acquiring rewards (Berridge and Robinson, 1998). According to this theory, the tonic dopamine neuron activity, which we observed in the present study, seems to be a good candidate for a neural substrate that regulates action selection in changing environments in which reward values fluctuate. That is, by tracking the changing values of future rewards, the tonic activity might enhance and suppress motivational vigor of actions to obtain the rewards on a moment-by-moment basis.

Conclusion

316 In conclusion, we demonstrated that, instead of phasic firing, tonic firing of dopamine
 317 neurons continuously tracked reward values changing moment-by-moment. This tonic
 318 activity was produced more largely by non-burst spikes than burst spikes that evoked
 319 the phasic dopamine neuron activity. Our findings expand the current knowledge on
 320 dopamine neuron signals by highlighting the alternating firing modes of these neurons,
 321 which effectively signal reward values in a given situation. Further investigations are
 322 called for to definitely determine whether the tonic activity of dopamine neurons
 323 represents state values, reward prediction errors or other reward-related variables.

Materials and Methods

Animals

Two adult macaque monkeys (*Macaca fuscata*; monkey H, female, 6.6 kg, 8 years old; monkey P, female, 6.8 kg, 7 years old) were used for the experiments. All procedures for animal care and experimentation were approved by the University of Tsukuba Animal Experiment Committee (permission number, 14-137), and were carried out in accordance with the guidelines described in *Guide for the Care and Use of Laboratory Animals* published by the Institute for Laboratory Animal Research.

Behavioral tasks

Behavioral tasks were controlled using MATLAB (Mathworks, MA, USA) with Psychtoolbox, a freely available toolbox for MATLAB. The monkeys sat in a primate chair facing a computer monitor in a sound-attenuated and electrically shielded room. Eye movements were monitored using an infrared eye-tracking system (EYE-TRAC 6; Applied Science Laboratories, MA, USA) with sampling at 240 Hz.

We designed a Pavlovian procedure with three different conditions (*Figure 1A-C*) and applied it to the two monkeys. In each condition, trials started with the presentation of a central fixation point ($0.8^\circ \times 0.8^\circ$). The monkey was required to fixate this point.

After a 400-ms fixation period, a bar stimulus ($2.9^\circ \times 9.4^\circ$) with red and green areas was presented as a conditioned stimulus (CS). The size of the green area indicated the amount (i.e., the value) of a liquid reward that the monkey would receive. The larger the green area became, the larger the associated reward value. In the first condition (value-increase condition, *Figure 1A*), the green area was minimal at the beginning and gradually increased ($3.8^\circ/\text{s}$), i.e., the reward value gradually increased (0.082 ml/s) (*Video 1*). The gradual increase randomly stopped within 2450 ms after the onset of the bar stimulus (uniform distribution from 0 to 2450 ms) so that the monkey was unable to precisely predict the reward value until the green area stopped increasing. In the second condition (value-decrease condition, *Figure 1B*), the green area was maximal at the beginning and gradually decreased ($3.8^\circ/\text{s}$), i.e., the reward value gradually decreased (0.082 ml/s) (*Video 2*). The gradual decrease randomly stopped within 2450 ms after the bar onset (uniform distribution from 0 to 2450 ms). In the third condition (value-fixed condition, *Figure 1C*), the size of the green area did not change and was instead fixed at the minimum, half, or maximum that predicted a 0.1-ml, 0.2-ml, or 0.3-ml reward, respectively (these reward amounts were the same as those indicated by the green area in the value-increase and value-decrease conditions). In addition to these CSs, a CS with a longitudinally halved green area was included, and this was randomly followed by

a 0.1- to 0.3-ml reward (i.e., the reward value was uncertain). These four CSs were presented with an equal probability (25% each). In each condition, the total time during which the CS was presented was fixed at 2850 ms, and the monkey was required to maintain the central fixation during this period. Immediately after the offset of the bar stimulus, the reward indicated by the green area was delivered simultaneously with a tone (1000 Hz). The delay between the CS onset and the reward delivery did not change across trials. Each condition consisted of a block of 50 trials, and we collected data by repeating the three conditions once or more for each neuron (*Figure 1D*). Trials were aborted immediately if the monkey 1) did not start the central fixation within 4000 ms after the onset of the fixation point, or 2) broke the central fixation during the initial 400-ms fixation period or the 2850-ms CS period (i.e., a continuous 3250-ms fixation was required). These error trials were signaled by a beep tone (100 Hz) and excluded from analyses.

We also used a choice task (*Figure 1E*). Each trial started with the presentation of a central fixation point ($0.8^{\circ} \times 0.8^{\circ}$), and the monkey was required to fixate this point. After a 400-ms fixation period, two bar stimuli with red and green areas ($2.9^{\circ} \times 9.4^{\circ}$) were presented on the right and left sides of the fixation point (eccentricity: 8.8°). The monkey was required to choose one of the bar stimuli by making a saccade immediately

after the presentation of the bar stimuli. These bar stimuli were identical to those used in the Pavlovian procedure except that the green area did not increase or decrease over time. The size of the green area indicated the value of the reward that the monkey would obtain by choosing that bar stimulus, and was randomly assigned to each bar stimulus (uniform distribution from 0.1 to 0.3 ml). Immediately after the monkey chose a bar stimulus, the other bar stimulus disappeared. Then, the reward indicated by the green area of the chosen bar stimulus was delivered simultaneously with a tone (1000 Hz). Trials were aborted immediately if the monkey 1) did not start the central fixation within 4000 ms after the onset of the fixation point, or 2) broke the central fixation during the 400-ms fixation period. These error trials were signaled by a beep tone (100 Hz) and excluded from analyses.

Single-unit recording

A plastic head holder and recording chamber were fixed to the skull under general anesthesia and sterile surgical conditions. The recording chamber was placed over the midline of the frontoparietal lobes, and aimed at the SNc and VTA in both hemispheres. The head holder and recording chamber were embedded in dental acrylic that covered the top of the skull and were firmly anchored to the skull by plastic screws.

Single-unit recordings were performed using tungsten electrodes with impedances of approximately 3.0 M Ω (Frederick Haer, ME, USA) that were introduced into the brain through a stainless-steel guide tube using an oil-driven micromanipulator (MO-97-S; Narishige, Tokyo, Japan). Recording sites were determined using a grid system, which allowed recordings at every 1 mm between penetrations. For a finer mapping of neurons, we also used a complementary grid that allowed electrode penetrations between the holes of the original grid.

Electrophysiological signals were amplified, band-pass filtered (200 Hz to 3 kHz; RZ5D, Tucker-Davis Technologies, FL, USA), and stored in a computer at the sampling rate of 24.4 kHz. Single-unit potentials were isolated using a window discrimination software (OpenEx, Tucker-Davis Technologies, FL, USA).

Localization and identification of dopamine neurons

We recorded single-unit activity from putative dopamine neurons in the SNc and VTA. To localize the recording regions, the monkeys underwent a magnetic resonance imaging (MRI) scan to determine the positions of the SNc and VTA (*Figure 2–figure supplement 1*). Putative dopamine neurons were identified based on their well-established electrophysiological signatures: a low background firing rate at around 5 Hz,

a broad spike potential in clear contrast to neighboring neurons with a high background firing rate in the substantia nigra pars reticulata (*Figure 2-figure supplement 2*), and a phasic excitation in response to free reward.

Data analysis

For null hypothesis testing, 95% confidence intervals ($p < 0.05$) were used to define statistical significance in all analyses.

To examine whether the monkeys accurately predicted the reward values according to the size of the green area (*Figure 1F*), the choice rate of the right bar stimulus was fitted by the following logistic function:

$$P = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 \times (V_{right} - V_{left})))}$$

where P indicates the choice rate of the right bar stimulus, V_{right} and V_{left} indicate the reward values obtained by choosing the right and left bar stimuli, respectively, and β_0 and β_1 indicate the coefficients determined by logistic regression.

To calculate spike density functions (SDFs), each spike was replaced by a Gaussian curve ($\sigma = 15$ ms).

To calculate SDFs aligned by the CS onset in the value-increase and value-decrease conditions (*Figure 2A, B, E, F*, and *Figure 5A*), spikes that occurred before the

reward value stopped increasing or decreasing were used.

To calculate averaged SDFs across neurons (*Figure 2E, F, Figure 3D and Figure 5A, B*), the baseline firing rate of each neuron was measured using a window from 500 to 0 ms before the fixation point onset, and was subtracted from the original firing rate of each neuron.

To quantify tonic activity changes in the value-increase and value-decrease conditions (*Figure 2C, D*), the slope of the regression line between firing rate and time was calculated for each neuron. First, a window from 650 to 2450 ms after the CS onset was divided into 200-ms bins, and the firing rate in each bin was measured. Then, the regression line between the firing rate in each bin and the time at the center of each bin was calculated. The calculation window from 650 to 2450 ms after the CS onset was determined to exclude the effect of phasic responses evoked immediately after the CS onset.

To examine phasic activity modulations evoked by the reward value in the value-fixed condition, the regression coefficient between firing rate and reward value was calculated for each neuron. The firing rate was measured using a window from 100 to 400 ms after the CS onset. The calculation window was determined such that the window includes the major part of the neural modulation in the averaged activity.

To examine phasic activity modulation evoked when the reward value stopped increasing or decreasing in the value-increase and value-decrease conditions (*Figure 3B, C*), the regression coefficient between firing rate and reward value was calculated for each neuron. First, all trials were divided into three groups based on the reward value (large: 0.23 to 0.3 ml, medium: 0.16 to 0.23 ml, small: 0.1 to 0.16 ml), and the firing rate was measured using a window from 150 to 500 ms after the stop onset for each trial group. Then, the regression coefficient between the firing rate and the mean reward value in each trial group was calculated. The calculation window was determined such that the window includes the major part of the neural modulation in the averaged activity.

To statistically test the difference in firing rate (original firing rate – baseline firing rate) between burst and non-burst spike firing (left in *Figure 5A*), a bootstrap procedure was applied. We first divided the calculation time window (650 to 2450 ms after the CS onset), which we used to detect the significantly positive regression slope of the 19 dopamine neurons (horizontal gray bar in the left column of *Figure 2E*), into initial (650 to 1250 ms), middle (1250 to 1850 ms), and late (1850 to 2450 ms) periods. The 19 dopamine neurons were randomly resampled with replacements to form a new bootstrap dataset that had the same number of neurons as the original dataset. Using the new dataset, we compared the firing rate (original firing rate – baseline firing rate) between

burst and non-burst spike firing for the initial, middle, and late periods. This random resampling and comparison process was repeated 1000 times. If the firing rate was larger for burst spike firing than for non-burst spike firing or vice versa in more than 975 repetitions, the difference in firing rate was regarded as significant ($p < 0.05$; bootstrap test with 1000 repetitions).

Acknowledgments

We thank K. Bunzui, T. Okano, and Y. Narita for animal care, and Dr. T. Kawai for supporting analyses. This research was supported by MEXT KAKENHI Grant Number JP16H06567 to M.M., and JST CREST Grant Number JPMJCR1853 to M.M. The monkeys used in the present study were provided by NBRP "Japanese Monkeys" through the National Bio Resource Project of the MEXT, Japan.

Competing interests

The authors declare no competing interests.

Author contributions

485 Y.W., O.T., and M.M. designed the research. Y.W. and O.T. performed the
486 experiments and analyzed the data. M.M. wrote the original draft. All authors discussed
487 the results and revised the manuscript. M.M. organized this project.

References

- Bayer HM, Glimcher PW. 2005. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**:129-141.
- Bayer HM, Lau B, Glimcher PW. 2007. Statistics of midbrain dopamine neuron spike trains in the awake primate. *J. Neurophysiol.* **98**:1428-1439.
- Berridge KC, Robinson TE. 1998. What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Res. Brain Res. Rev.* **28**:309-369.
- Brischoux F, Chakraborty S, Brierley DI, Ungless MA. 2009. Phasic excitation of dopamine neurons in ventral vta by noxious stimuli. *Proc. Natl. Acad. Sci. USA* **106**:4894-4899.
- Cagniard B, Balsam PD, Brunner D, Zhuang X. 2006. Mice with chronically elevated dopamine exhibit enhanced motivation, but not learning, for a food reward. *Neuropsychopharmacology* **31**:1362-1370.
- Chang CY, Esber GR, Marrero-Garcia Y, Yau HJ, Bonci A, Schoenbaum G. 2016. Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. *Nat. Neurosci.* **19**:111-116.
- Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N. 2012. Neuron-type-specific signals

506 for reward and punishment in the ventral tegmental area. *Nature* **482**:85-88.

507 da Silva JA, Tecuapetla F, Paixao V, Costa RM. 2018. Dopamine neuron activity before

508 action initiation gates and invigorates future movements. *Nature* **554**:244-248.

509 de Lafuente V, Romo R. 2011. Dopamine neurons code subjective sensory experience

510 and uncertainty of perceptual decisions. *Proc. Natl. Acad. Sci. USA* **108**:19767-

511 19771.

512 Doya K. 2002. Metalearning and neuromodulation. *Neural Netw.* **15**:495-506.

513 Engelhard B, Finkelstein J, Cox J, Fleming W, Jang HJ, Ornelas S, Koay SA, Thiberge

514 SY, Daw ND, Tank DW, Witten IB. 2019. Specialized coding of sensory, motor

515 and cognitive variables in vta dopamine neurons. *Nature* **570**:509-513.

516 Fiorillo CD, Tobler PN, Schultz W. 2003. Discrete coding of reward probability and

517 uncertainty by dopamine neurons. *Science* **299**:1898-1902.

518 Grace AA, Bunney BS. 1984a. The control of firing pattern in nigral dopamine neurons:

519 Burst firing. *J. Neurosci.* **4**:2877-2890.

520 Grace AA, Bunney BS. 1984b. The control of firing pattern in nigral dopamine neurons:

521 Single spike firing. *J. Neurosci.* **4**:2866-2876.

522 Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Vander Weele CM,

523 Kennedy RT, Aragona BJ, Berke JD. 2016. Mesolimbic dopamine signals the

524 value of work. *Nat. Neurosci.* **19**:117-126.

525 Howard CD, Li H, Geddes CE, Jin X. 2017. Dynamic nigrostriatal dopamine biases action
526 selection. *Neuron* **93**:1436-1450.

527 Howe MW, Dombeck DA. 2016. Rapid signalling in distinct dopaminergic axons during
528 locomotion and reward. *Nature* **535**:505-510.

529 Howe MW, Tierney PL, Sandberg SG, Phillips PE, Graybiel AM. 2013. Prolonged
530 dopamine signalling in striatum signals proximity and value of distant rewards.
531 *Nature* **500**:575-579.

532 Kawagoe R, Takikawa Y, Hikosaka O. 2004. Reward-predicting activity of dopamine and
533 caudate neurons--a possible mechanism of motivational control of saccadic eye
534 movement. *J. Neurophysiol.* **91**:1013-1024.

535 Maes EJP, Sharpe MJ, Usypchuk AA, Lozzi M, Chang CY, Gardner MPH, Schoenbaum
536 G, Iordanova MD. 2020. Causal evidence supporting the proposal that
537 dopamine transients function as temporal difference prediction errors. *Nat.*
538 *Neurosci.* **23**:176-178.

539 Matsumoto M, Hikosaka O. 2009. Two types of dopamine neuron distinctly convey
540 positive and negative motivational signals. *Nature* **459**:837-841.

541 Matsumoto M, Takada M. 2013. Distinct representations of cognitive and motivational

542 signals in midbrain dopamine neurons. *Neuron* **79**:1011-1024.

543 Menegas W, Akiti K, Amo R, Uchida N, Watabe-Uchida M. 2018. Dopamine neurons
544 projecting to the posterior striatum reinforce avoidance of threatening stimuli.
545 *Nat. Neurosci.* **21**:1421-1430.

546 Mohebi A, Pettibone JR, Hamid AA, Wong JT, Vinson LT, Patriarchi T, Tian L, Kennedy
547 RT, Berke JD. 2019. Dissociable dopamine dynamics for learning and
548 motivation. *Nature* **570**:65-70.

549 Montague PR, Dayan P, Sejnowski TJ. 1996. A framework for mesencephalic dopamine
550 systems based on predictive hebbian learning. *J. Neurosci.* **16**:1936-1947.

551 Morrens J, Aydin C, Janse van Rensburg A, Esquivelzeta Rabell J, Haesler S. 2020.
552 Cue-evoked dopamine promotes conditioned responding during learning.
553 *Neuron* **106**:142-153.

554 Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H. 2004. Coincident but distinct
555 messages of midbrain dopamine and striatal tonically active neurons. *Neuron*
556 **43**:133-143.

557 Niv Y. 2007. Cost, benefit, tonic, phasic: What do response rates tell us about dopamine
558 and motivation? *Ann. NY Acad. Sci.* **1104**:357-376.

559 Nomoto K, Schultz W, Watanabe T, Sakagami M. 2010. Temporally extended dopamine

560 responses to perceptually demanding reward-predictive stimuli. *J. Neurosci.*
561 **30**:10692-10702.

562 Ogasawara T, Nejime M, Takada M, Matsumoto M. 2018. Primate nigrostriatal dopamine
563 system regulates saccadic response inhibition. *Neuron* **100**:1513-1526 e1514.

564 Roesch MR, Calu DJ, Schoenbaum G. 2007. Dopamine neurons encode the better
565 option in rats deciding between differently delayed or sized rewards. *Nat.*
566 *Neurosci.* **10**:1615-1624.

567 Satoh T, Nakai S, Sato T, Kimura M. 2003. Correlated coding of motivation and outcome
568 of decision by dopamine neurons. *J. Neurosci.* **23**:9913-9923.

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

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

572 Stauffer WR, Lak A, Yang A, Borel M, Paulsen O, Boyden ES, Schultz W. 2016.
573 Dopamine neuron-specific optogenetic stimulation in rhesus macaques. *Cell*
574 **166**:1564-1571.

575 Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. 2013. A causal
576 link between prediction errors, dopamine neurons and learning. *Nat. Neurosci.*
577 **16**:966-973.

- 578 Tobler PN, Fiorillo CD, Schultz W. 2005. Adaptive coding of reward value by dopamine
579 neurons. *Science* **307**:1642-1645.
- 580 Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K. 2009.
581 Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning.
582 *Science* **324**:1080-1084.
- 583 Yun M, Kawai T, Nejime M, Yamada H, Matsumoto M. 2020. Signal dynamics of midbrain
584 dopamine neurons during economic decision-making in monkeys. *Sci. Adv.*
585 **6**:eaba4962.

Figures

Figure 1. Pavlovian procedure in which the value of a cued reward gradually changed.

(A-C) Value-increase (A), value-decrease (B) and value-fixed (C) conditions in the Pavlovian procedure. **(D)** Schedule of recording session. Each condition consisted of a block of 50 trials and was repeated once or more. **(E)** Choice task. **(F)** Choice rate of the

594 right bar stimulus as a function of the value difference between the right and left bar
595 stimuli in monkey H (n = 85 sessions) (black) and monkey P (n = 68 sessions) (gray).
596 Error bars indicate SEM.

Figure 2. Tonic activity of dopamine neurons tracking gradually changing reward values.

(A and B) Activity of two example dopamine neurons in the value-increase (left), value-decrease (middle) and value-fixed (right) conditions. Rasters and spike density functions (SDFs) are aligned by the CS onset. Left: Red and orange rasters indicate spikes occurring before and after, respectively, the reward value stopped increasing. Gray plots indicate the time at which the reward value stopped changing. The spikes occurring

605 before the stop were used to calculate the SDF. The thick gray line indicates the
606 regression line between firing rate and time. The bottom graph indicates the reward value
607 as a function of time. Middle: Blue and cyan rasters indicate spikes occurring before and
608 after, respectively, the reward value stopped decreasing. Other conventions are as in the
609 left panel. Right: Black (top), dark gray, light gray, and black (bottom) rasters indicate
610 spikes in trials in which the reward value was large, medium, small, and uncertain,
611 respectively. Black (solid), dark gray, light gray, and black (dotted) curves indicate the
612 SDFs in these trials. **(C and D)** Regression slope between firing rate and time in the
613 value-increase (C) and value-decrease (D) conditions ($n = 99$ neurons). Red and gray
614 bars indicate neurons with significantly positive and negative slopes, respectively, in the
615 value-increase condition (C), while blue and gray bars indicate neurons with a
616 significantly negative and positive slope, respectively, in the value-decrease condition
617 (D) ($p < 0.05$; linear regression analysis). Arrowheads indicate the mean regression
618 slopes. Single and double asterisks indicate a significant deviation from zero ($p < 0.05$
619 and $p < 0.01$, respectively; Wilcoxon signed-rank test). **(E and F)** Averaged SDFs of the
620 19 dopamine neurons with a significantly positive regression slope in the value-increase
621 condition (E) and those of the 15 dopamine neurons with a significantly negative
622 regression slope in the value-decrease condition (F). Left: Red and blue curves indicate

623 the SDFs in the value-increase and value-decrease conditions, respectively. Right: Black
 624 (solid), dark gray, light gray, and black (dotted) curves indicate the SDFs in trials in which
 625 the reward value was large, medium, small, and uncertain, respectively. Horizontal gray
 626 bars indicate the time windows used for analyses.

627 **Figure supplement 1.** Localization of recording sites.

628 **Figure supplement 2.** Spike waveforms of dopamine neurons.

629 **Figure supplement 3.** Relationship between the regression slopes in the value-increase
 630 and value-decrease conditions.

Figure 3. Phasic activity of dopamine neurons evoked when the reward value stopped changing.

(A) Activity of an example dopamine neuron. Rasters and SDFs are aligned by the time at which the reward value stopped increasing or decreasing in the value-increase (left) and value-decrease (right) conditions, respectively. Left: Dark red, red, and light red indicate the activity in trials in which the stopping of the value increase occurred at early (small reward: 0.1 to 0.16 ml), middle (medium reward: 0.16 to 0.23 ml) and late (large reward: 0.23 to 0.3 ml) times, respectively. Right: Dark blue, blue, and light blue indicate the activity in trials in which the stopping of the value decrease occurred at early (large reward: 0.23 to 0.3 ml), middle (medium reward: 0.16 to 0.23 ml) and late (small reward: 0.1 to 0.16 ml) times, respectively.

0.1 to 0.16 ml) times, respectively. **(B and C)** Regression coefficient between firing rate and reward value in the value-increase (B) and value-decrease (C) conditions (n = 99 neurons). Conventions are as in *Figure 2C, D*. **(D)** Averaged SDFs of the 37 dopamine neurons with a significantly positive regression coefficient in the value-increase condition (left) and those of the 27 dopamine neurons with a significantly positive regression coefficient in the value-decrease condition (right). Horizontal gray bars indicate the time windows used for analyses. Conventions are as in *Figure 3A*.

Figure 4. Relationship between the tonic and phasic dopamine neuron activities.

(A) Regression slope between firing rate and time for the tonic activity increase in the value-increase condition (ordinate) and regression coefficient between firing rate and reward value in the value-fixed condition (abscissa) ($n = 99$ neurons). **(B)** Regression slope between firing rate and time for the tonic activity decrease in the value-decrease condition (ordinate) and regression coefficient between firing rate and reward value in the value-fixed condition (abscissa) ($n = 99$ neurons). **(C)** Regression slope between firing rate and time for the tonic activity decrease in the value-decrease condition (ordinate) and regression coefficient between firing rate and reward value for the stop-aligned phasic activity decrease in the value-increase condition (abscissa) ($n = 99$ neurons). **(D)** Regression slope between firing rate and time for the tonic activity increase

661 in the value-increase condition (ordinate) and regression coefficient between firing rate
 662 and reward value for the stop-aligned phasic activity increase in the value-decrease
 663 condition (abscissa) (n = 99 neurons). Green and orange dots indicate neurons with a
 664 significant difference from zero in the ordinate and abscissa, respectively ($p < 0.05$; linear
 665 regression analysis). Purple dots indicate neurons with significance for both.

Figure 5. Phasic and tonic activities of dopamine neurons produced by burst and non-burst spikes.

(A) Averaged SDFs of the 19 dopamine neurons with a tonic increase in activity in the value-increase condition. Left: SDFs in the value-increase condition. Right: SDFs in the value-fixed condition. The SDFs are aligned by the CS onset, and shown for burst (dotted curve) and non-burst (solid curve) spikes. Horizontal white bars indicate the initial, middle, and late periods to calculate the neural modulation between burst and non-burst spike firing. Single and double asterisks above the horizontal bars indicate periods during which the firing rate (original firing rate – baseline firing rate) was significantly larger for non-burst spike firing than for burst spike firing ($p < 0.05$ and $p < 0.01$, respectively; bootstrap test with 1000 repetitions). The SDFs in the value-fixed condition were calculated using large reward trials. **(B)** Averaged SDFs of the 27 dopamine neurons with a phasic activity that was modulated by reward values when the values stopped decreasing in the value-decrease condition. The SDFs in large reward trials are shown,

681 and aligned by the time at which the reward value stopped decreasing. Other
682 conventions are as in *Figure 5A*.

Figure supplements

Figure 2-figure supplement 1. Localization of recording sites.

Images obtained by MRI scans for monkey H (left) and monkey P (right) are shown.

Vertical red lines indicate representative recording tracks. The approximate anteroposterior distance (millimeters) from the interaural line is shown in the top-left corner of each panel.

Figure 2-figure supplement 2. Spike waveforms of dopamine neurons.

Distributions of the spike widths of dopamine neurons (left, $n = 99$) and non-dopamine neurons in the substantia nigra pars reticulata (right, $n = 12$). Red bars indicate dopamine neurons with a significantly positive regression slope between firing rate and time in the value-increase condition, while blue bars indicate dopamine neurons with a significantly negative regression slope in the value-decrease condition ($p < 0.05$; linear regression analysis). Purple bars indicate dopamine neurons with both of them, and open bars indicate dopamine neurons with neither of them. The spike widths of the dopamine neurons with a significantly positive regression slope in the value-increase condition, those with a significantly negative regression slope in the value-decrease condition, and those with neither of them were not significantly different (positive versus negative: $p = 0.77$; positive versus no significance: $p = 0.051$; negative versus no significance: $p = 0.058$; Wilcoxon rank-sum test). Spike shapes of example dopamine and non-dopamine

705 neurons are shown in the top-right corner of each panel. The two vertical lines show how
706 the spike duration was measured.

707

708 **Figure 2–figure supplement 3. Relationship between the regression slopes in the**
 709 **value-increase and value-decrease conditions.**

710 Regression slope between firing rate and time in the value-increase condition (ordinate)
 711 and that in the value-decrease condition (abscissa) (n = 99 neurons). Green and orange
 712 dots indicate dopamine neurons with a significant regression slope in the value-increase
 713 and value-decrease conditions, respectively ($p < 0.05$; linear regression analysis). Purple
 714 dots indicate neurons with a significant regression slope in both conditions.

Rich media files

Video 1. Bar stimuli in the value-increase condition of the Pavlovian procedure.

The size of the green area indicated the amount of a liquid reward. The green area was minimal at the beginning and gradually increased. The gradual increase randomly stopped within 2450 ms after the onset of the bar stimulus.

Video 2. Bar stimuli in the value-decrease condition of the Pavlovian procedure.

The green area was maximal at the beginning and gradually decreased. The gradual increase randomly stopped within 2450 ms after the onset of the bar stimulus.