nature neuroscience

The temporal precision of reward prediction in dopamine neurons

Christopher D Fiorillo[1], William T Newsome[1] & Wolfram Schultz[2]

Midbrain dopamine neurons are activated when reward is greater than predicted, and this error signal could teach target neurons both the value of reward and when it will occur. We used the dopamine error signal to measure how the expectation of reward was distributed over time. Animals were trained with fixed-duration intervals of 1–16 s between conditioned stimulus onset and reward. In contrast to the weak responses that have been observed after short intervals (1–2 s), activations to reward increased steeply and linearly with the logarithm of the interval. Results with varied stimulus-reward intervals suggest that the neural expectation was substantial after just half an interval had elapsed. Thus, the neural expectation of reward in these experiments was not highly precise and the precision declined sharply with interval duration. The neural precision of expectation appeared to be at least qualitatively similar to the precision of anticipatory licking behavior.

To select the most advantageous behavioral outputs, the nervous system must recognize and exploit spatial and temporal patterns in its sensory inputs. Although spatial processing has a straightforward anatomical correlate, very little is known about how the nervous system processes time. Interval timing in the range of seconds to hours is known to depend on the basal ganglia and, like other functions of the basal ganglia, is strongly dependent on dopamine innervation from the ventral midbrain[1–5]. Several models of timing propose a central role for midbrain dopamine neurons[5–14]. A number of these are reinforcement learning models that rely on the reward prediction error signal of midbrain dopamine neurons to teach target neurons to predict when a reward event will occur (as well as its reward value)[7–14]. We investigated the effect of reward timing on the responses of dopamine neurons.

Evidence that dopamine neurons encode a reward prediction error comes primarily from electrophysiological recordings[15–24]. Dopamine neurons are strongly activated when a reward, such as a drop of juice, is repeatedly delivered with a long and variable inter-reward interval. However, when a reward consistently follows a conditioned stimulus or an operant response after a fixed delay of 1–2 s, dopamine neurons show little or no activation. When reward is expected, but does not occur, then the neurons' firing rate is suppressed below its baseline level shortly after the time that reward is usually delivered, even in the absence of any external cue[15,16,21]. Furthermore, when the expectation of reward is systematically varied, the activation by reward is graded, declining roughly in proportion to reward expectation[18,21,24]. Thus, by measuring the dopamine error signal while manipulating the timing of reward, we should be able to estimate the temporal precision of the neural expectation. In principle, the neural expectation could be quite precise, rising abruptly to its peak near the end of a fixed interval, or could be less precise, rising slowly from the start of the interval. Furthermore, temporal precision at the behavioral level is known to decline as interval duration increases, and dopamine responses could therefore be sensitive to interval duration.

RESULTS

Behavior

Two macaque monkeys underwent Pavlovian conditioning (**Fig. 1a**). Each of four visual stimuli, presented on randomly interleaved trials, was associated with a particular fixed delay to liquid reward, which varied from 1–16 s depending on the conditioned stimulus. Each monkey was trained for at least 5 d and 600 presentations of each conditioned stimulus to establish stable conditioned behavior before the start of physiological recordings.

Anticipatory licking served as a behavioral measure of the monkeys' reward expectation. Sweetened liquid was delivered from a spout just in front of the mouth. If the monkey's tongue or lips were not already extended toward the spout at the time of the liquid delivery, a portion of the liquid would fall unconsumed. After training, both monkeys showed anticipatory licking only during stimulus-reward intervals. Licking typically started about halfway through the stimulus-reward interval and, once it had begun, lasted until reward was delivered (**Fig. 1b**). Both monkeys licked earlier and on a greater proportion of trials when the interval was shorter (**Fig. 1c**), although monkey B licked on a smaller proportion of trials than monkey A (see Methods).

After we completed the physiological recordings, we trained the animals on a variant of the same task called the peak-interval procedure[1]. The same conditioned stimuli were used, but the conditioned stimulus remained on for three times its usual duration and no reward was delivered on one quarter of pseudo-randomly chosen trials. The average licking behavior on these probe trials increased until the usual

[1]Department of Neurobiology, Fairchild Building, D209, 299 Campus Drive West, Stanford University, Stanford, California 94305-5125, USA. [2]Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK. Correspondence should be addressed to C.D.F. (chris@monkeybiz.stanford.edu).

Received 17 March; accepted 3 June; published online 27 July 2008; doi:10.1038/nn.2159

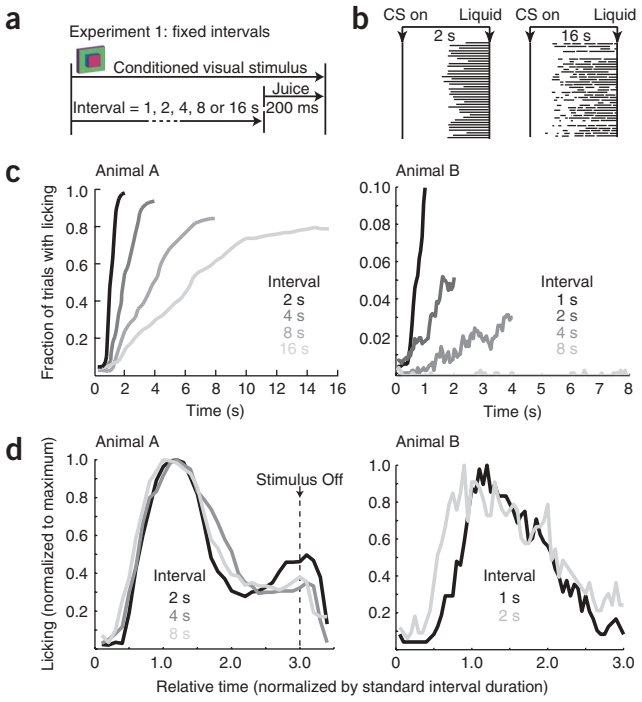

Figure 1 Timing of anticipatory licking behavior (Experiment 1). (**a**) The task design used Pavlovian delay conditioning. Each of four visual icons, presented on pseudo-randomly interleaved trials, was paired with a different delay from conditioned stimulus onset to juice onset. Receipt of the full volume of juice required licking at the time of juice delivery. (**b**) Each horizontal line represents the time of licking on an individual trial during 2- and 16-s intervals. The time was normalized to interval duration. CS, conditioned stimulus. (**c**) The fraction of all trials with licking increased as the interval elapsed. Conditioned stimulus onset occurred at time zero. Licking is shown up until the time of juice delivery. (**d**) Licking behavior on unrewarded probe trials in the peak interval procedure. Time (*x* axis) was scaled by the standard stimulus-reward interval. The overlap of data for different stimulus-reward intervals demonstrates that the timing of licking scales with interval duration, consistent with Weber's Law, except on the rising phase in animal B.

time of reward and then declined, suggesting that the monkeys' subjective expectation usually reached its peak very near the end of the standard interval (**Fig. 1d**). However, the licking was spread over a relatively long period of time, with the amount of licking on probe trials remaining over half its maximal value for at least as long as the standard interval.

A hallmark of interval timing in the range of seconds to minutes or longer is that the temporal uncertainty, or spread in the expectation, scales linearly with the interval duration[1,4,25]. This phenomenon is often called the 'scalar' property of timing, and it is analogous to Weber's law for the perception of stimulus intensity. When the time axes were scaled by the standard interval and the average amount of licking was normalized by its maximum, the licking superimposed across stimulus-reward intervals (**Fig. 1d**). Thus the licking behavior of our monkeys appeared to show scalar timing. The only exception was licking on the rising phase in monkey B (see Methods).

Observation of licking behavior led us to several conclusions. First, the fact that licking was restricted to the latter half of stimulus-reward intervals and was sometimes entirely absent for longer intervals (depriving the animals of reward) suggests that the cost of licking was high enough that the animals had an incentive for precise timing in this task. Nonetheless, even after hundreds of trials of training on fixed intervals, licking typically began after only about half the interval had elapsed. This suggests substantial temporal uncertainty (low precision). However, the temporal uncertainty cannot be readily quantified, as licking is unlikely to directly reflect reward expectation. Presumably the monkey licks when its reward expectation at a given moment exceeds a threshold that is determined by the cost of licking. This would explain why both the latency and average amount of licking depend on how much liquid is expected[23]. Thus, an absence of licking does not necessarily reflect an absence of expectation, and the expectation presumably rises both before and after the onset of licking. Finally, temporal precision declined as the duration of the interval was increased (**Fig. 1c,d**). Thus, the momentary expectation was lower during longer stimulus-reward intervals in comparison with shorter intervals. The patterns of timed behavior in our monkeys were at least qualitatively similar to those of humans and other animals[1,4,25].

Physiology

We individually recorded single dopamine neurons in the substantia nigra and ventral tegmental area (see Methods). Juice was delivered without any predictive cue in a separate block of trials with a long and variable inter-reward interval (a uniform distribution of 6–16 s in monkey A and 2–16 s in monkey B). As in our previous work[18,23], reward delivered in this context was less predictable than in any other condition that we studied. The relatively large response to this 'unpredicted' reward served as a standard to which other reward responses in the same neuron could be compared.

We carried out three distinct experiments. The first examined the effect of fixed stimulus-reward intervals of different durations (**Fig. 2**). On the basis of the behavior, the spread of temporal expectation appeared to increase (precision decreases) with stimulus-reward interval (**Fig. 1**). Thus, the momentary expectation was weaker for longer interval durations, and this appeared to be reflected in the larger reward responses of dopamine neurons (**Fig. 2**). The second experiment examined the response to reward that was delivered earlier or later than usual on a small number of probe trials (**Fig. 3**). The third experiment examined responses to reward that was delivered after a stimulus-reward interval that varied in duration from trial to trial (**Figs. 4** and **5**). These latter two experiments provided a measure of the temporal spread or precision of expectation. Dopamine reward responses in these tasks were found to be only weakly sensitive to the precise timing of reward, suggesting low precision.

Experiment 1: fixed intervals of differing durations

Neurons were recorded in the Pavlovian task described above (**Fig. 1a–c**). As previously shown, when a conditioned stimulus preceded reward by a short interval of 1 or 2 s, the reward elicited little or no activation, whereas the conditioned stimulus elicited a substantial activation. As the stimulus-reward interval increased, the response to the conditioned stimulus diminished (**Fig. 2a,b** and **Supplementary Fig. 1** online). This finding supports the proposal of dopamine-mediated reinforcement learning models that the reward value that drives dopamine neurons is best defined as the discounted sum of future rewards[7–14]. According to these models, a conditioned stimulus associated with a longer delay to liquid reward acquires a lesser reward value, as the value of the liquid that it predicts is discounted.

In contrast to conditioned stimulus responses, reward responses increased with the stimulus-reward interval (**Fig. 2c,d** and **Supplementary Fig. 1**). Linear regression analyses showed that the mean population response to reward in each monkey increased in proportion

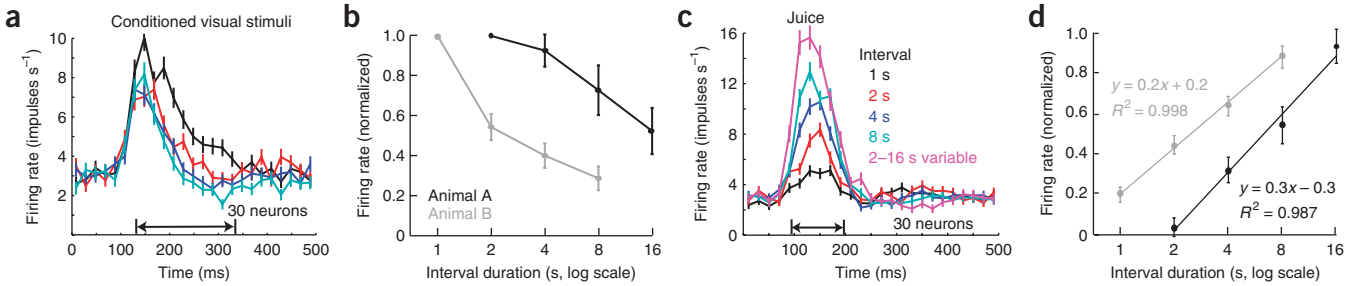

Figure 2 Dopamine neurons are sensitive to interval duration (Experiment 1). (**a**) Population histograms (mean ± s.e.m.) of firing rates from 30 neurons in animal B following the onset of conditioned stimuli. Neurons were more strongly activated by stimuli associated with a shorter delay before reward. Color code is shown in **c**. Bin size = 20 ms. The time period indicated by the double-headed arrow was chosen to quantify the conditioned stimulus responses of all neurons in animal B, as shown in **b**. Histograms of firing rate are shown in the same format in all figures. (**b**) Mean population responses (± s.e.m., 11 neurons in animal A, 30 in animal B) to each conditioned stimulus as a function of stimulus-reward interval. Prior to averaging across neurons, responses were normalized to the response following the conditioned stimulus that was associated with the shortest interval (see Methods). (**c**) Population histograms of firing rates from 30 neurons following juice onset in animal B. (**d**) Mean population responses to juice delivery as a function of stimulus-reward interval (26–34 neurons in animal A (black), 30 in animal B (gray)). Prior to averaging across neurons, responses were normalized to the response to juice delivery that followed a long and variable interval (see Methods). For each animal, responses were a linear function of interval duration on a logarithmic scale, as indicated by the R^2 values.

to the logarithm of the interval ($R^2 > 0.98$, $P < 0.01$; **Fig. 2d**). Among individual neurons, 38 of 40 had a positive slope, of which 16 were significantly positive ($P < 0.05$; **Supplementary Fig. 2** and **Supplementary Results** online). Thus, in contrast with previous observations with short fixed intervals, rewards delivered after fixed intervals that were 2–4-fold longer caused a substantial activation of dopamine neurons. In fact, after a 16-s interval, the reward response was as large as that following 'unpredicted' reward (**Fig. 2d**). This is presumably because the temporal spread of expectation increased and the momentary expectation of reward decreased with interval duration, as suggested by the lesser amount of licking behavior that we observed during longer intervals (**Fig. 1c**).

Experiment 2: probe trails with early or late reward

To investigate the extent to which the expectation is spread over time, we carried out a second set of experiments in which we trained each animal with a fixed stimulus-reward interval of 2.0 s for 500–2,000 trials. During subsequent physiological recordings, we delivered reward after a delay of 1.0 s on a small number of probe trials (**Fig. 3a**). Early reward caused a greater activation than did reward at its usual time (**Fig. 3b,c**), confirming a previous study[16]. However, there was much less activation to early reward than to 'unpredicted' reward, suggesting that the neural expectation was relatively strong even after just half of the 2-s interval had elapsed. We observed similar responses when reward was delivered at 1.5, 2.5 or 3.0 s (**Supplementary Fig. 3** and **Supplementary Results** online). We found no physiological evidence that the reward expectation had been substantially altered by exposure to the probe trials, and even if it were, the alteration would probably not be large enough to account for the weak activation to early reward that we observed (**Fig. 3**; see **Supplementary Fig. 4** and **Supplementary Results** online).

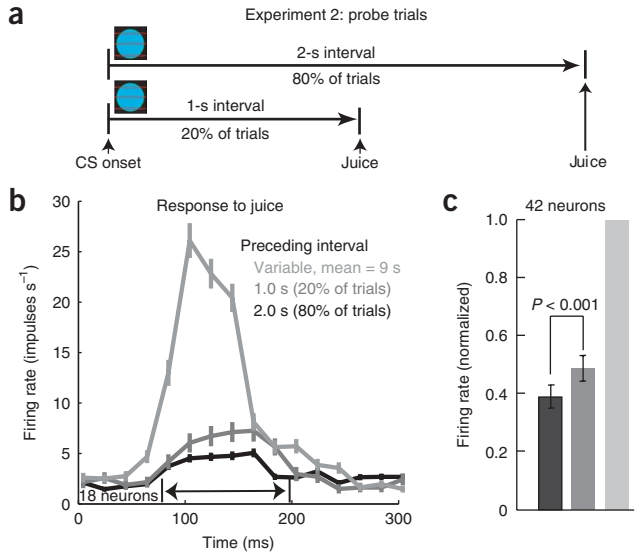

Figure 3 Response of dopamine neurons to juice delivered earlier or later than usual (Experiment 2). (**a**) Schematic illustrating the design of Experiment 2 with probe trials. After extensive training in which a conditioned stimulus was followed by juice after 2.0 s on all trials, juice was delivered after 1.0 s on 20% of pseudo-randomly selected probe trials during physiological recordings. (**b**) Population histograms (animal B) of responses to juice when it was delivered at its usual time of 2.0 s, at 1.0 s on probe trials, or unpredictably after a long and variable interval in the absence of an explicit conditioned stimulus. (**c**) Mean (± s.e.m.) responses averaged across all neurons in both animals, measured at the time period indicated by the double-headed arrow shown in **b**. Prior to averaging, firing rates in each neuron were divided by the response to 'unpredicted' juice in the same neuron. Gray scale is the same as **b**.

Experiment 3: variable intervals

As an additional test of the temporal precision of the neural expectation, we examined the effect of varying a stimulus-reward interval from trial to trial (**Fig. 4a**). The distribution of intervals was flat and ranged between 1.0 and 3.0 s (or 0.5 and 3.5 s; **Fig. 5a**). The animal's expectation during this variable interval should be spread over a duration of at least 2 s (as a result of averaging across past intervals of differing durations; for example, see ref. 26). If the expectation associated with each observed interval is strong for only a brief period near the end of the interval (high precision), then introduction of variable intervals should substantially weaken the momentary reward expectation, resulting in large dopamine responses. In contrast with this high-precision hypothesis, we found that dopamine neurons were

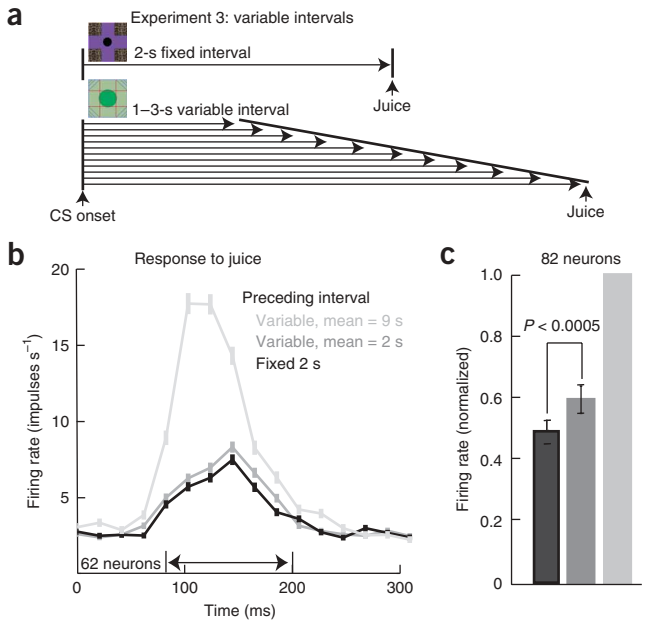

Figure 4 Response of dopamine neurons to juice delivered following a stimulus-reward interval that varied across trials (Experiment 3). (**a**) Schematic illustrating the design of Experiment 3. Onset of one conditioned stimulus was always followed by juice after a fixed interval of 2.0 s. Onset of a second conditioned stimulus was followed by juice at any time between 1.0 and 3.0 s. (**b**) Population histograms (animal B) of responses to juice delivered after a fixed 2.0-s interval, a variable interval with a mean of 2.0 s, or a long and variable interval. We tested 32 of the 62 neurons recorded in animal B with a 0.5–3.5-s interval rather than a 1–3-s interval. All responses following the variable interval conditioned stimulus have been averaged together, regardless of whether a particular interval was long or short. **Figure 5** shows responses as a function of the actual preceding interval. (**c**) Mean (± s.e.m.) responses averaged across all neurons in both animals, measured at the time period indicated by the double-headed arrow shown in **b**. Prior to averaging, the firing rates in each neuron were divided by the response to 'unpredicted' juice of the same neuron. Gray scale is the same as **b**.

only slightly (but significantly) more activated by reward delivered after the variable interval than after a fixed 2-s interval ($P < 0.005$ in each monkey, paired t tests; **Fig. 4b**,**c**), suggesting that trial-to-trial variations in interval duration of ±50% did not strongly alter the expectation. Thus, the neural expectation appeared to be relatively high during both fixed and variable intervals of short duration, as evidenced by the much larger reward responses that we observed with the standard long and variable intervals (**Fig. 4b**,**c**).

When the reward responses in the same experiment were examined as a function of the preceding interval, responses were observed to be larger on trials in which reward was delivered after a shorter interval (**Fig. 5b**), suggesting that the expectation grew stronger as the interval elapsed in the absence of reward. Because reward was delivered by 3.0 s on every trial, the absence of reward at one moment during the elapsing interval may increase the expectation that reward will occur in the next moment. The contingency over time between no reward and reward is described by the hazard function. It has been shown that the behavior of monkeys and the firing rate of neurons in parietal cortex track elapsed time in a manner consistent with the hazard function[27]. Although the expectation apparently grew during the elapsing interval, late rewards nonetheless caused activation of dopamine neurons above their baseline firing rate (compare the activation rates in **Fig. 5b** with the baseline rates in **Fig. 4b**). Similarly, delivery of conditioned stimuli and rewards following longer than average intertrial intervals also elicited activations (data not shown). These observations appear to contradict a previously proposed model of reward timing[14], which predicts that rewards delivered later than usual should suppress the activity of dopamine neurons.

If the expectation of reward grows over time during an elapsing interval, then the activity of dopamine neurons may be increasingly suppressed as the stimulus-reward interval elapses in the absence of reward. Indeed, the firing rate gradually declined as the variable interval elapsed ($P < 0.05$ in each monkey; see Methods and **Fig. 5c**). The slight gradual decline in firing rate could reflect negative prediction errors that are small and growing over time as reward expectation increases. We also found evidence for a similar effect during short fixed stimulus-reward intervals. For a fixed 1-s stimulus-reward interval, the firing rate in a population of 30 neurons was suppressed from a baseline of 3.00 ± 0.08 impluses per s (measured in the 0.5-s period before conditioned stimulus onset) to 2.70 ± 0.07 in the 0.5-s period before reward delivery ($P < 0.01$, paried t-test). A similar suppression of activity was seen for a fixed 2-s interval ($P < 0.05$, 62 neurons; **Fig. 5c**), but not for longer intervals (**Fig. 5f**). Such an effect may be negligible for longer fixed intervals, as the momentary expectation would be weaker.

In the experiment described above, the distribution of stimulus-reward intervals was flat, meaning that all of the intervals were equally probable. Previous work has shown that the behavior of monkeys and the activity of single cortical neurons can also adapt to reflect knowledge of unimodal and bimodal interval distributions[27,28]. An exponential distribution is particularly interesting to consider, as its hazard function is flat. Thus the expectation in the present moment is independent of what happened in the preceding moment. We trained monkey B with a stimulus-reward interval of 1 s plus an exponential distribution with a mean of 1 s (**Fig. 5d**). Responses of dopamine neurons were larger following earlier rewards ($P < 0.05$, linear regression; **Fig. 5e**), similar to our results with flat distributions (**Fig. 5b**). In addition, the firing rate tended to decline gradually as the interval elapsed with no reward delivery, although this trend did not reach statistical significance ($P = 0.22$; **Fig. 5c**). Thus, neural activity appeared not to reflect knowledge of the unique time-independent property of the exponential distribution. Despite considerable training, the monkey may not have been able to learn that neighboring points in time were statistically independent. We found this interesting, as one might have suspected that statistical independence is readily learned or even that it represents the system's default assumption. However, exponential distributions of intervals between reward events may be virtually nonexistent in the natural world, and a system that is designed to exploit temporal correlations may not be capable of leaning that there is no correlation in some rare cases.

A model of timing in dopamine neurons

We attempted to develop a mathematical description of the temporal expectation that is consistent with the behavioral and neural data and is able to explain why reward responses increase in proportion to the logarithm of the stimulus-reward interval. For the case of familiar fixed intervals, we wanted to describe how the animal's subjective reward expectation is distributed over future time, starting just after the conditioned stimulus has appeared and been identified by the animal. For simplicity, we assume that the expectation is reasonably described by a probability distribution, in which case the integral over future time must be one (even though reward can never be certain to be delivered in the future). Neither the behavioral nor neural data

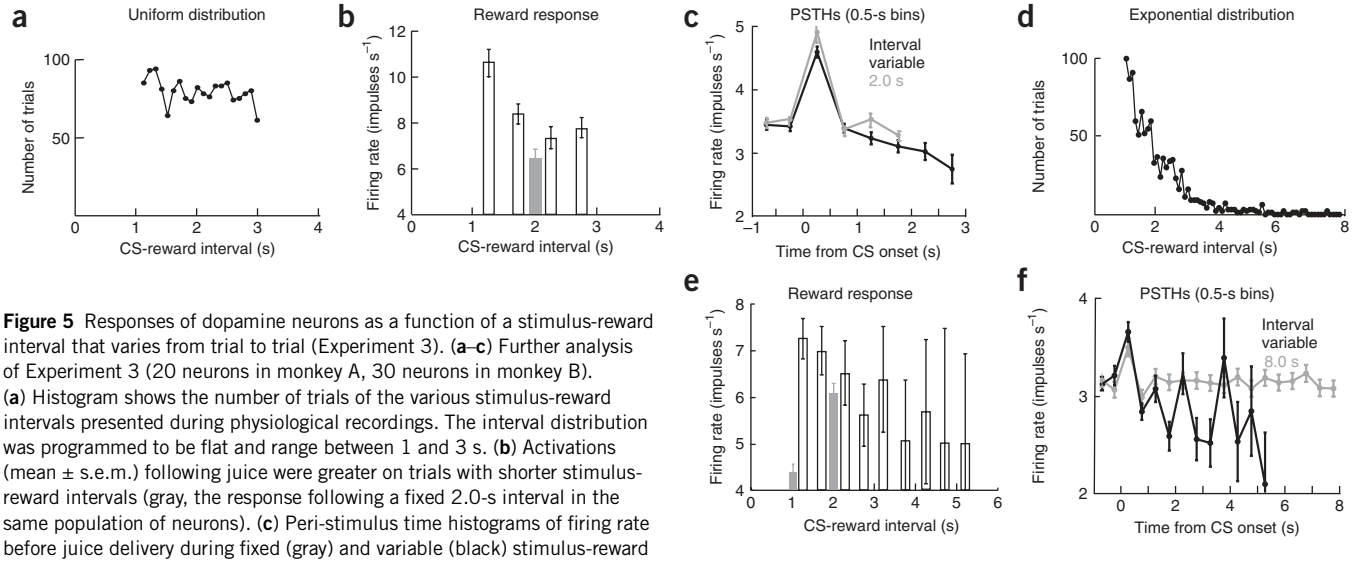

Figure 5 Responses of dopamine neurons as a function of a stimulus-reward interval that varies from trial to trial (Experiment 3). (**a**–**c**) Further analysis of Experiment 3 (20 neurons in monkey A, 30 neurons in monkey B). (**a**) Histogram shows the number of trials of the various stimulus-reward intervals presented during physiological recordings. The interval distribution was programmed to be flat and range between 1 and 3 s. (**b**) Activations (mean ± s.e.m.) following juice were greater on trials with shorter stimulus-reward intervals (gray, the response following a fixed 2.0-s interval in the same population of neurons). (**c**) Peri-stimulus time histograms of firing rate before juice delivery during fixed (gray) and variable (black) stimulus-reward intervals. Following the initial phasic activation in response to conditioned stimulus onset, firing rate gradually declined as the variable interval elapsed, as indicated by linear regression analyses. Error bars (s.e.m.) are larger for longer intervals, as fewer long duration trials were recorded. We obtained virtually identical results to **c** and **b** from a separate population of 32 neurons recorded in monkey B with a variable stimulus-reward interval of 0.5 to 3.5 s (data not shown). (**d**–**f**) We recorded from 26 neurons in monkey B during presentation of an exponential distribution of stimulus-reward intervals. (**d**) Histogram shows the number of trials of the various stimulus-reward intervals when the interval distribution was programmed to be exponential with a mean of 2.0 s. (**e**) As for the flat distribution, activations following juice were greater on trials with shorter stimulus-reward intervals (gray, response following fixed intervals of 1 and 2 s in the same population of neurons). The increase in the size of error bars with interval duration (**e**,**f**) is the result of the small number of long duration trials recorded (**d**). (**f**) Peri-stimulus time histograms of firing rate before juice delivery during fixed (gray) and variable (black) stimulus-reward intervals. In contrast to the significant time-dependent modulation of firing rate in **b**, **c** and **e**, linear regression analysis did not reveal a significant decline in firing rate during the stimulus-reward interval.

place strong constraints on the shape or width of the distribution (see Discussion). However, for a variety of different distributions that are consistent with Weber's law, the peak expectation would decline in proportion to the inverse of the interval duration (**Fig. 6a** and Methods). If this is correct, then the present experiments have systematically varied reward expectation over an eightfold range, which is twice the range of previous studies[18,21]. Indeed, at the longest interval tested (16 s), the activation of dopamine neurons was as large as our standard 'unpredicted' reward (**Fig. 2d**), which evoked the largest responses seen in this and previous studies[18,23]. Over this relatively large range of expected values and firing rates, the firing rate following reward was a linear function of the logarithm of the stimulus-reward interval (**Fig. 2d**). We considered what form of prediction error could account for these results.

The conventional prediction error of temporal difference learning algorithms depends on the difference between the actual and expected reward value at a given moment in time. Assuming that the expectation declines as the inverse of the interval duration, the error at the end of a familiar stimulus-reward interval of duration T would be

$$\text{Error} = R - \frac{R}{T} \qquad (1)$$

where R is reward magnitude (which is constant in the present experiments). The conventional prediction error would therefore be a highly nonlinear function of interval duration (**Fig. 6b**) and of the log of duration (not shown), and thus it appears not to provide an accurate description of the responses of dopamine neurons when interval duration is varied.

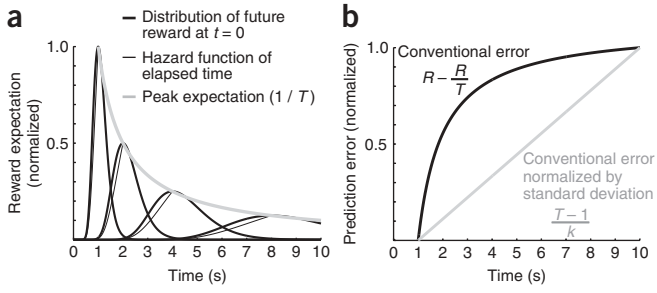

Figure 6 A model of interval timing and the dopamine error signal that could account for the data of **Figure 2d**. (**a**) The subjective distributions of reward expectation over future time (immediately following the perceptual identification of the conditioned stimulus at time 0) after training on fixed stimulus-reward intervals of 1, 2, 4 and 8 s (thick black curves). The distributions shown here are log normal (equation (3) with $\sigma = 0.25$). However, neither the width nor the precise shape of the distribution was important for explaining the mean population responses of Experiment 1 (**Fig. 2d**). Thin black curves represent hazard functions. The peak of the expectation (which coincides with the time of reward delivery) is inversely related to interval duration, as illustrated by the gray curve. (**b**) The conventionally defined reward prediction error (equation (1)) is shown in black as a function of the duration of the preceding stimulus-reward interval. Dividing the conventional prediction error by the s.d. of the estimate of reward magnitude (which was assumed to vary in proportion to the expectation, as suggested by Weber's law, equation (2)) yielded the gray line. This normalized prediction error is a linear function of interval duration and could account for the data from Experiment 1 (**Fig. 2d**). This assumes that the firing rate corresponds to the log of the error, as suggested by the Weber-Fechner law when applied to reward magnitude, and thus **Figure 2d** can be understood as a log-log plot.

Even disregarding the present data, two modifications to the conventionally defined prediction error are probably necessary to accurately describe the error signal of dopamine neurons. First, the Weber-Fechner law presumably applies to the perception of reward magnitude[26,29] and its encoding by dopamine neurons. Second, it has been shown that dopamine neurons are sensitive to reward uncertainty[18] and that the error response of dopamine neurons appears to be scaled or normalized by a measure of uncertainty such as the s.d.[23]. Thus, the prediction at each moment in time would correspond to a distribution of potential reward magnitudes and the error signal would be the difference between actual and expected value in units of s.d. This would be an efficient means of representing reward value, analogous to the common practice in statistics of measuring a value in z scores or standard deviates from the mean.

To incorporate these two modifications, we first needed to divide the conventional prediction error of equation (1) by the s.d. of the estimate of reward magnitude. We do not have a measure of the animal's uncertainty about reward magnitude, but Weber's law suggests that the uncertainty will vary in proportion to the expected value (R/T). The s.d. would therefore be equal to $k\frac{R}{T}$, where k is the coefficient of variation in reward magnitude (the ratio of the s.d. to the expected (mean) reward magnitude). Thus, when reward is delivered at the end of an interval of duration T, the modified prediction error would be

$$\text{Error}_{\text{normalized}} = \frac{R - \frac{R}{T}}{k\frac{R}{T}} = \frac{T-1}{k}. \qquad (2)$$

If the error is normalized in this way, it would grow as a linear function of interval duration (**Fig. 6b**). Thus the normalized error could account for the data from Experiment 1 (**Fig. 2d**) if we assume that the firing rate of dopamine neurons depends not on the actual difference between reward magnitude and expected magnitude, but rather on the logarithmically scaled perception of the difference, as suggested by the Weber-Fechner law as applied to reward magnitude[26,29]. The graph of mean population responses (**Fig. 2d**) could then be thought of as a log-log plot, with firing rate corresponding to the logarithm of the normalized prediction error. The linearity in the log-log plot (**Fig. 2d**) would mirror that shown in **Figure 6b** with linear axes. In summary, the mean population responses of Experiment 1 (**Fig. 2d**) may be explained by the confluence of three hypotheses, each of which is already supported by independent evidence. First, the peak expectation of reward is inversely related to interval duration, as required for a variety of distributions of future reward that are consistent with the scalar timing property. Second, the dopamine error signal is the conventionally defined error divided by a measure of uncertainty such as s.d.[23]. Finally, the activity of dopamine neurons encodes a logarithmic transformation of the error signal, as implied by the Weber-Fechner law when applied to reward magnitude.

DISCUSSION

We found that dopamine neurons were highly sensitive to the duration of a stimulus-reward interval (**Fig. 2**), but were only weakly sensitive to the precise timing of reward following a conditioned stimulus (**Figs. 3–5**). Similar to the behavioral expectation, as measured by Pavlovian-conditioned licking behavior, the neural expectation appeared to be substantial after just half of a familiar interval had elapsed. Thus, our results suggest that the expectation of reward in these Pavlovian tasks was not of high temporal precision at either the neural or behavioral level and that temporal precision declined sharply as interval duration increased.

Receipt of reward in the Pavlovian task studied here did not require behaviorally precise timing, and thus another task may have revealed more precise timing. However, our monkeys were extensively trained and their behavior demonstrated that they had both knowledge of stimulus-reward intervals and an incentive for precise timing. Furthermore, it is generally believed that the marked flexibility of interval timing, functioning over a large range of interval durations from seconds to hours, comes at the cost of precision[4]. For example, a 10–30% difference between intervals is required for even moderately reliable discrimination[1]. At the neuronal level, a lack of high precision might be inferred from the expectation-related delay-period activity of many sensorimotor neurons (for example, see refs. 27,28,30–33), including dopamine neurons[18]. Firing rates typically begin to gradually increase (or decrease) shortly after the first event and they continue until the second event occurs, suggesting that the neuronal expectation is characterized by high temporal uncertainty. We came to a similar conclusion by observing the phasic prediction error of dopamine neurons following the second event.

In addition to the gradual delay-period activation, and the phasic reward activation studied here, the inhibition of dopamine neurons following the omission of predicted reward also relates to the precision of timing. A previous study[16] reported that when juice was expected at 1.0 s after a manual response, the inhibition became substantial at 1.1 s on average across individual neurons. This inhibition must be internally timed and could reflect the activation of some neurons in which interval timing is more precise than in dopamine neurons. However, we found evidence in the summed activity of a population of neurons that a slight suppression of firing rate below baseline was already present between 0.5 and 1.0 s after the start of a 1.0-s fixed interval. Regardless of the inhibition's temporal precision, our results address its involvement in the computation of the prediction error. Anatomically specific computational models have suggested that the inhibition of dopamine neurons is timed to cancel the excitation, thereby producing the error signal[7,10–12,34]. However, we found that the excitatory effect of reward can be greatly diminished even at times when the absence of reward does not lead to any measurable inhibition (**Figs. 3–5**). This suggests that the cancellation of excitation to predictable rewards may be largely complete before the excitatory reward signal reaches dopamine neurons and that the inhibition of dopamine neurons may only provide the 'finishing touch' to a computation that is largely performed upstream of dopamine neurons.

Both the early onset of licking behavior (**Fig. 1**) and the weak neural responses to early reward (**Figs. 3–5**) suggest that reward predictions are characterized by substantial temporal uncertainty. However, the weak activations to early reward may not result solely from the limited temporal precision of the system. First, there may be a time-independent component to reward expectation resulting from a simple stimulus-reward association, as suggested by the late, stimulus-driven licking behavior (**Fig. 1d**). A second contribution is suggested by reinforcement learning models. According to these models, reward value at a given moment in time is defined as the expected sum of current and future rewards, where the values of future rewards are discounted according to their anticipated delay. In the present experiments, each presentation of a conditioned stimulus was accompanied by exactly one drop of juice. If the animal has learned this rule, then the reward value of early juice in experiments 2 and 3 is diminished, as it signals the absence of juice in the subsequent seconds. This aspect of predictive learning algorithms captures the common sense notion that the exact timing of a predicted reward should not matter much. Thus, the potential dependence of the dopamine error signal on both current and future rewards may complicate our effort to use it as an assay of the animal's subjective reward expectation. Consideration of each of the two factors discussed above

suggest that the temporal precision of reward prediction is probably greater than one might first think on the basis of only the weak responses to early reward in experiments 2 and 3. Nonetheless, these data clearly suggest the limited potential of dopamine neurons to signal the precise time of reward events.

A large and influential class of interval timing models relies on the existence of discrete elements that vary in their kinetic properties across a spectrum (for example, see ref. 35). Just as each of a set of neurons or synapses may be dedicated to a distinct region of space, spectral models propose that discrete neural elements are dedicated to distinct periods of the past (or equivalently, the future). When an external stimulus occurs, it triggers the full spectrum of 'time-stamped' memory traces. The goal of the system is to select those memory traces that best predict when a second reward-related event will occur. As learning is only needed when the second event is not accurately predicted, it has been proposed that the prediction error of dopamine neurons drives the selection process[8,9,11–13]. Climbing fibers from the inferior olivary nucleus may have an analogous role in training the cerebellum[2,36]. In the absence of neural data on the temporal uncertainty in prediction, most models have not addressed the issue. Instead they have assumed, for simplicity, that the spectral components have arbitrarily precise timing that does not decline as the duration of the interval increases (but see ref. 12). Such models therefore imply that the activation of dopamine neurons by reward may be quite sensitive to variation in the timing of reward, but should be absent following reward at the end of a familiar fixed interval, regardless of its duration. Our results indicate just the opposite. A more realistic future model might include a spectrum of timing components that are similar to those depicted in **Figure 6a**, in which the expectation of each component would be substantially spread out over time and its spread would grow in proportion to the maximum likelihood estimate of the interval.

Although the ability for precise timing would appear to be advantageous, its utility may be limited under typical conditions. The statistics of natural reward events have not been characterized, but one might expect that intervals that require internal timing tend to be quite short. This is because relatively stereotyped sequences of conditioned stimuli may be common, effectively providing the animal with an external 'clock'. If so, then internal timing may usually be needed only for the short intervening intervals between stimuli (or 'ticks' of the clock). In addition, the repeated occurrence of very precisely timed intervals on which we trained animals may be highly unusual in more natural settings. Relatively imprecise timing could therefore suffice if the majority of important, recurring intervals tend to be short and variable. If, for whatever reason, a system lacks the capacity for precise internal timing of long intervals, then the accurate prediction of reward must rely on identifying external conditioned stimuli that precede reward by short intervals. The activation of dopamine neurons after long stimulus-reward intervals would be useful for identifying shorter stimulus-reward intervals (assuming that other predictive stimuli are present in the environment). Thus, our results suggest that the dopamine error signal would be well suited for a world in which the majority of intervals between reward events tend to be short and variable.

METHODS

Animals. We studied two adult female macaques, monkey A at the University of Fribourg (*Macaca fascicularis*, 3.3 kg) and monkey B at Stanford University (*Macaca mulatta*, 10.5 kg). All experimental procedures complied with guidelines established by the Swiss Animal Protection Law and the US National Institutes of Health, and were overseen locally by the Fribourg Cantonal Veterinary Office and the Stanford University Animal Care and Use Committee.

Experimental design. We used 0.2 ml of flavored, sweetened liquid ('juice') as a reward, which was delivered over 120 ms to monkey A and 200 ms to monkey B from a spout immediately in front of the animal's mouth. Each experiment included 1–6 conditioned stimuli, each presented on pseudo-randomly interleaved trials. Conditioned stimuli were arbitrary visual icons 5 deg in diameter (such as those shown in **Figs. 1a, 3a** and **4a**), each presented at a distinct location on a computer monitor for monkey A or in the same central location for monkey B. The intertrial interval from juice delivery on one trial to conditioned stimulus onset on the next trial was variable and ranged with a flat distribution between 6–16 s in monkey A and 2–6 s in monkey B. For each experiment, 1 week of pre-training was followed by about 4 weeks of recording. Approximately 100–200 trials of each stimulus were presented each day. Recordings began only after a pre-training period of about 5 d and 600 presentations of each conditioned stimulus, accompanied by the emergence of discriminative conditioned licking responses.

Behavior. Licking (and eye position in monkey B) was monitored as described in the **Supplementary Methods** online. Monkey B licked on fewer trials than monkey A and did not show any measurable licking during the longest interval tested (8 s; **Fig. 1c**). The most likely explanation of the difference in behavior between the two animals is that, although the internal expectation of reward rose gradually in both animals during the delay interval, monkey B had a higher threshold to begin licking. The lesser amount of licking was not simply the result of a general lack of attention to the stimuli, as monkey B consistently looked toward the (predictable) position of the conditioned stimulus (before conditioned stimulus onset on 32% of trials and in the first 300 ms after conditioned stimulus onset on 89% of trials). Although the behavior of monkey B does not tell us whether or not the 8-s interval had been learned, the smaller activation to reward after an 8-s interval (compared with 'unpredicted' reward) suggests that learning had taken place (**Fig. 2c,d**). The lesser amount of licking could have been in part because monkey B was accustomed to several-fold shorter intertrial intervals than monkey A (see above), and therefore stimulus-reward intervals may have been more aversive to monkey B (or similarly, stimulus-reward associations may have been weaker).

The licking in monkey B on the rising phase of the peak interval procedure did not superimpose after scaling time by interval duration (**Fig. 1d**). This may have been because the 1-s interval was short enough that the onset of licking was influenced by limits on reaction time or because the scalar timing property breaks down for intervals less than about 0.5–1.0 s[1,2].

Recording dopamine neurons. Single-unit recordings were carried out as described previously[15]. Dopamine neurons were distinguished by their discharge characteristics. See the **Supplementary Methods** for additional information.

Data analysis. Similar to previous physiological studies, we found that the population of dopamine neurons was quite homogeneous (for example, see refs. 15–24,37). Therefore, we have summarized the data by averaging across the entire population of recorded neurons. Because responses were relatively stereotyped in their time course, the same standard windows were used across all neurons to calculate firing rate following conditioned stimulus or juice onset (see **Supplementary Methods** for additional information).

We normalized firing rates in **Figure 2** and **Supplementary Figure 3** by measuring the last 500 ms of the intertrial interval before conditioned stimulus onset to determine a neuron's baseline firing rate. The baseline rate was subtracted from the rate measured in the standard window following juice or conditioned stimulus onset. The resulting value was then divided by the difference between baseline activity and the response in the same neuron to 'unpredicted' juice (delivered following a long and variable intertrial interval, or to the response to the conditioned stimulus associated with the shorted interval in the case of conditioned stimulus responses). Normalized values tended to lie between 0 and 1, but were occasionally outside of this range.

Linear regression analyses were performed on the normalized mean population responses of Experiment 1 (**Fig. 2d**) and on the unnormalized firing rates in Experiment 3 (**Fig. 5c,e,f**). In the former case, neuronal responses to reward

(mean responses in single cells or for the population) were regressed on the logarithm (base 2) of the preceding stimulus-reward interval. In the latter case, firing rates (as shown for the population averages in **Fig. 5**) were calculated in 0.5-s bins for every trial in every cell (1,100–1,900 trials) and these rates were then regressed on the time since conditioned stimulus onset. For peri-stimulus time histograms (**Fig. 5c,f**), firing rates before and up to 500 ms after conditioned stimulus onset were excluded. Because trials with short stimulus-reward intervals did not contribute data to the later bins, this approach weighted the early time bins more heavily than later bins.

Model. We wanted to find the subjective distribution of reward expectation as a function of time after conditioned stimulus onset in the case of familiar fixed intervals. The shape of the distribution is not well constrained by the data, but it could be normally distributed on a log scale[38]. Thus, the expectation at the beginning of the interval $E(t)_{t=0}$ would be distributed over future time t as:

$$E(t)_{t=0} = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(\frac{-(\ln t - \mu)^2}{2\sigma^2}\right) \quad (3)$$

For a log-normal distribution to be consistent with scalar timing (Weber's law), the s.d. on a log scale (σ) would need to be a constant (independent of interval duration)[38]. The peak subjective expectation should occur approximately at the end of a fixed interval, which in the case of a log normal distribution would occur when time t is equal to the mode ($\exp(\mu - \sigma^2)$) in equation (3) (**Fig. 6a**). Substituting $\exp(\mu - \sigma^2)$ for t in equation (3) shows that the exponential term of equation (3) would be independent of interval duration and the peak expectation would decline in proportion to the inverse of the interval duration (**Fig. 6a**). Because our concern here is restricted to how the peak expectation scales with interval duration, neither the spread of the distribution nor its precise shape is important. Indeed, the inverse relationship of peak expectation to interval duration is also true for certain other distributions that are consistent with scalar timing, such as Rayleigh and normal distributions with a constant coefficient of variation.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

This work was supported by grants from the Human Frontiers Science Program (C.D.F.), the Howard Hughes Medical Institute (W.T.N.), the US National Institutes of Health (EY 05603, W.T.N.), the Swiss National Science Funds (W.S.) and the Wellcome Trust (W.S.).

AUTHOR CONTRIBUTIONS

C.D.F. conducted the experiments, analyzed the data and developed the mathematical model of dopamine responses. C.D.F. and W.S. designed the experiments. C.D.F. wrote the manuscript with feedback from W.T.N. and W.S.

Published online at http://www.nature.com/natureneuroscience/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Buhusi, C.V. & Meck, W.H. What makes us tick? Functional and neural mechanisms of interval timing. *Nat. Rev. Neurosci.* **6**, 755–765 (2005).
2. Mauk, M.D. & Buonomano, D.V. The neural basis of temporal processing. *Annu. Rev. Neurosci.* **27**, 307–340 (2004).
3. Rao, S.M., Mayer, A.R. & Harrington, D.L. The evolution of brain activation during temporal processing. *Nat. Neurosci.* **4**, 317–323 (2001).
4. Gibbon, J., Malapani, C., Dale, C.L. & Gallistel, C.R. Toward a neurobiology of temporal cognition: advances and challenges. *Curr. Opin. Neurobiol.* **7**, 170–184 (1997).
5. Meck, W.H. Neuropharmacology of timing and time perception. *Brain Res. Cogn. Brain Res.* **3**, 227–242 (1996).
6. Matell, M.S. & Meck, W.H. Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Brain Res. Cogn. Brain Res.* **21**, 139–170 (2004).
7. Houk, J., Adams, J. & Barto, A. A model of how the basal ganglia generate and use neural signals that predict reinforcement. *Models of Information Processing in the Basal Ganglia* (eds Houk, J., Davis, J. & Beiser, D.) 249–270 (MIT Press, Cambridge, Massachusetts, 1995).
8. Montague, P.R., Dayan, P. & Sejnowski, T.J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–1947 (1996).
9. Schultz, W., Dayan, P. & Montague, R.R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
10. Berns, G.S. & Sejnowski, T.J. A computational model of how the basal ganglia produce sequences. *J. Cogn. Neurosci.* **10**, 108–121 (1998).
11. Brown, J., Bullock, D. & Grossberg, S. How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *J. Neurosci.* **19**, 10502–10511 (1999).
12. Contreras-Vidal, J.L. & Schultz, W. A predictive reinforcement model of dopamine neurons for learning approach behavior. *J. Comput. Neurosci.* **6**, 191–214 (1999).
13. Suri, R.E. & Schultz, W. A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience* **91**, 871–890 (1999).
14. Daw, N.D., Courville, A.C. & Touretzky, D.S. Representation and timing in theories of the dopamine system. *Neural Comput.* **18**, 1637–1677 (2006).
15. Schultz, W., Apicella, P. & Ljungberg, T. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.* **13**, 900–913 (1993).
16. Hollerman, J.R. & Schultz, W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* **1**, 304–309 (1998).
17. Schultz, W. The predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
18. Fiorillo, C.D., Tobler, P.N. & Schultz, W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* **299**, 1898–1902 (2003).
19. Satoh, T., Nakai, S., Sato, T. & Kimura, M. Correlated coding of motivation and outcome of decision by dopamine neurons. *J. Neurosci.* **23**, 9913–9923 (2003).
20. Nakahara, H., Itho, H., Kawagoe, R., Takikawa, Y. & Hikosaka, O. Dopamine neurons can represent context-dependent prediction error. *Neuron* **41**, 269–280 (2004).
21. Morris, G., Arkadir, D., Nevet, A., Vaadia, E. & Bergman, H. Coincident, but distinct, messages of midbrain dopamine and striatal tonically active neurons. *Neuron* **43**, 133–143 (2004).
22. Pan, W.X., Schmidt, R., Wickens, J.R. & Hyland, B.I. Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *J. Neurosci.* **25**, 6235–6242 (2005).
23. Tobler, P.N., Fiorillo, C.D. & Schultz, W. Adaptive coding of reward value by dopamine neurons. *Science* **307**, 1642–1645 (2005).
24. Bayer, H.M. & Glimcher, P.W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**, 129–141 (2005).
25. Rakitin, B.C. *et al.* Scalar expectancy theory and peak-interval timing in humans. *J. Exp. Psychol. Anim. Behav. Process.* **24**, 15–33 (1998).
26. Kacelnik, A. & Brito e Abreu, F. Risky choice and Weber's Law. *J. Theor. Biol.* **194**, 289–298 (1998).
27. Janssen, P. & Shadlen, M.N. A representation of the hazard rate of elapsed time in macaque area LIP. *Nat. Neurosci.* **8**, 234–241 (2005).
28. Ghose, G.M. & Maunsell, J.H.R. Attentional modulation in visual cortex depends on task-timing. *Nature* **419**, 616–620 (2002).
29. Bateson, M. & Kacelnik, A. Accuracy of memory for amount in the foraging starling, *Sturnus vulgaris*. *Anim. Behav.* **50**, 431–443 (1995).
30. Komura, Y. *et al.* Retrospective and prospective coding for predicted reward in the sensory thalamus. *Nature* **412**, 546–549 (2001).
31. Brody, C.D., Hernandez, A., Zanos, A. & Romo, R. Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cereb. Cortex* **13**, 1196–1207 (2003).
32. Leon, M.I. & Shadlen, M.N. Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron* **38**, 317–327 (2003).
33. Renoult, L.Roux. S. & Riehle, A. Time is a rubberband: neuronal activity in monkey motor cortex in relation to time estimation. *Eur. J. Neurosci.* **23**, 3098–3108 (2006).
34. O'Reilly, R.C., Frank, M.J., Hazy, T.E. & Watz, B. PVLV: the primary value and learned value Pavlovian learning algorithm. *Behav. Neurosci.* **121**, 31–49 (2007).
35. Grossberg, S. & Schmajuk, N.A. Neural dynamics of adaptive timing and temporal discrimination during associative learning. *Neural Netw.* **2**, 79–102 (1989).
36. Medina, J.F., Nores, W.L. & Mauk, M.D. Inhibition of climbing fibers is a signal for the extinction of conditioned eyelid responses. *Nature* **416**, 330–333 (2002).
37. Schultz, W. & Romo, R. Responses of nigrostriatal dopamine neurons to high-intensity somatosensory stimulation in the anesthetized monkey. *J. Neurophysiol.* **57**, 201–217 (1987).
38. Nieder, A. & Miller, E.K. Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* **37**, 149–157 (2003).