Behavioral/Cognitive

Phasic Dopamine Release in the Rat Nucleus Accumbens Symmetrically Encodes a Reward Prediction Error Term

Andrew S. Hart,[1]* Robb B. Rutledge,[2]* Paul W. Glimcher,[2] and Paul E. M. Phillips[1]

[1]Department of Psychiatry and Behavioral Sciences, Department of Pharmacology and Graduate Program in Neurobiology and Behavior, University of Washington, Seattle, Washington 98195, and [2]Center for Neural Science, New York University, New York, New York 10003

Making predictions about the rewards associated with environmental stimuli and updating those predictions through feedback is an essential aspect of adaptive behavior. Theorists have argued that dopamine encodes a reward prediction error (RPE) signal that is used in such a reinforcement learning process. Recent work with fMRI has demonstrated that the BOLD signal in dopaminergic target areas meets both necessary and sufficient conditions of an axiomatic model of the RPE hypothesis. However, there has been no direct evidence that dopamine release itself also meets necessary and sufficient criteria for encoding an RPE signal. Further, the fact that dopamine neurons have low tonic firing rates that yield a limited dynamic range for encoding negative RPEs has led to significant debate about whether positive and negative prediction errors are encoded on a similar scale. To address both of these issues, we used fast-scan cyclic voltammetry to measure reward-evoked dopamine release at carbon fiber electrodes chronically implanted in the nucleus accumbens core of rats trained on a probabilistic decision-making task. We demonstrate that dopamine concentrations transmit a bidirectional RPE signal with symmetrical encoding of positive and negative RPEs. Our findings strengthen the case that changes in dopamine concentration alone are sufficient to encode the full range of RPEs necessary for reinforcement learning.

Introduction

Hypotheses on the function of the neurotransmitter dopamine provide a link among theories of reinforcement learning, rational choice, and motivated behavior (Schultz et al., 1997, Morris et al., 2006, Phillips et al., 2007, Gan et al., 2010, Glimcher, 2011). A large number of studies have now correlated dopaminergic neural activity with the reward prediction error (RPE) term in temporal difference models of learning (Sutton, 1988, Montague et al., 1996, Schultz et al., 1997, Waelti et al., 2001, Flagel et al., 2011, Steinberg et al., 2013). Dopaminergic firing rates have been found to covary with this RPE term both qualitatively (Schultz et al., 1997) and quantitatively (Bayer and Glimcher, 2005). However, these studies have failed to differentiate putative dopaminergic RPE signals from other related signals such as salience or surprise. In response to this uncertainty, Caplin and Dean (2007) developed a mathematically rigorous definition of RPE signals using three axioms that are the necessary and sufficient conditions that must be fulfilled by any signal in order for it to be equivalent to an RPE representation.

Rutledge et al. (2010) were the first to use the axioms to test a signal in the brain for RPE equivalence. They found that during a task in which humans received probabilistic rewards, the BOLD signal in the nucleus accumbens meets necessary and sufficient criteria for encoding RPEs. Based upon these observations, they concluded that, in their task, an RPE-based learning algorithm could be driven by a neural signal in the nucleus accumbens. Although these findings confirmed that the nucleus accumbens carries an RPE-equivalent signal, they left some issues unresolved. The nature of the relationship between the BOLD signal and the extracellular dopamine concentration in a brain region is not fully understood, so it is unclear whether the RPE-equivalent signal in the nucleus accumbens is a reflection of dopamine release. This point is a critical shortcoming because dopamine is the centerpiece of the most prevalent RPE-based learning theories, such as temporal difference reinforcement learning (Schultz et al., 1997).

Here, we report electrochemical measurements of dopamine release in the nucleus accumbens core (NAc) of rats performing a simple set of behavioral tasks and demonstrate that these dopamine signals unambiguously encode an RPE representation in our experiments. Our quantitative measurements also indicate that dopamine release, unlike dopamine neuron firing rates, can encode positive and negative reward prediction errors on a symmetric scale. Although the present study does not address the roles of other neurotransmitter systems in reinforcement learning, our findings support the theoretical claim that bidirectional changes in dopamine concentration in the NAc could sufficiently encode all positive and negative RPEs in a reinforcement-learning task.

Materials and Methods

Animals and surgery. The University of Washington Institutional Animal Care and Use Committee approved all animal procedures used in this experiment. We implanted eight male Sprague Dawley rats with chronic carbon fiber electrodes (Clark et al., 2010) targeted bilaterally to the NAc

Received June 11, 2013; revised Oct. 18, 2013; accepted Nov. 15, 2013.

Author contributions: A.S.H., R.B.R., P.W.G., and P.E.M.P. designed research; A.S.H. and R.B.R. performed research; A.S.H., R.B.R., and P.E.M.P. analyzed data; A.S.H., R.B.R., P.W.G., and P.E.M.P. wrote the paper.

This work was supported by the National Institutes of Health (Grant R01-NS054775 to P.W.G. and Grants R01-MH079292 and R01-DA027858 to P.E.M.P.). We thank Scott Ng-Evans and Christina Akers for technical support.

The authors declare no competing financial interests.

*A.S.H. and R.B.R. contributed equally to this work.

Correspondence should be addressed to Paul E. M. Phillips, Department of Psychiatry and Behavioral Sciences, University of Washington, 1959 NE Pacific Street, Box 356560, Seattle, WA 98195-6560. E-mail: pemp@uw. Edu.

DOI:10.1523/JNEUROSCI.2489-13.2014

 0270-6474/14/340698-08$15.00/0

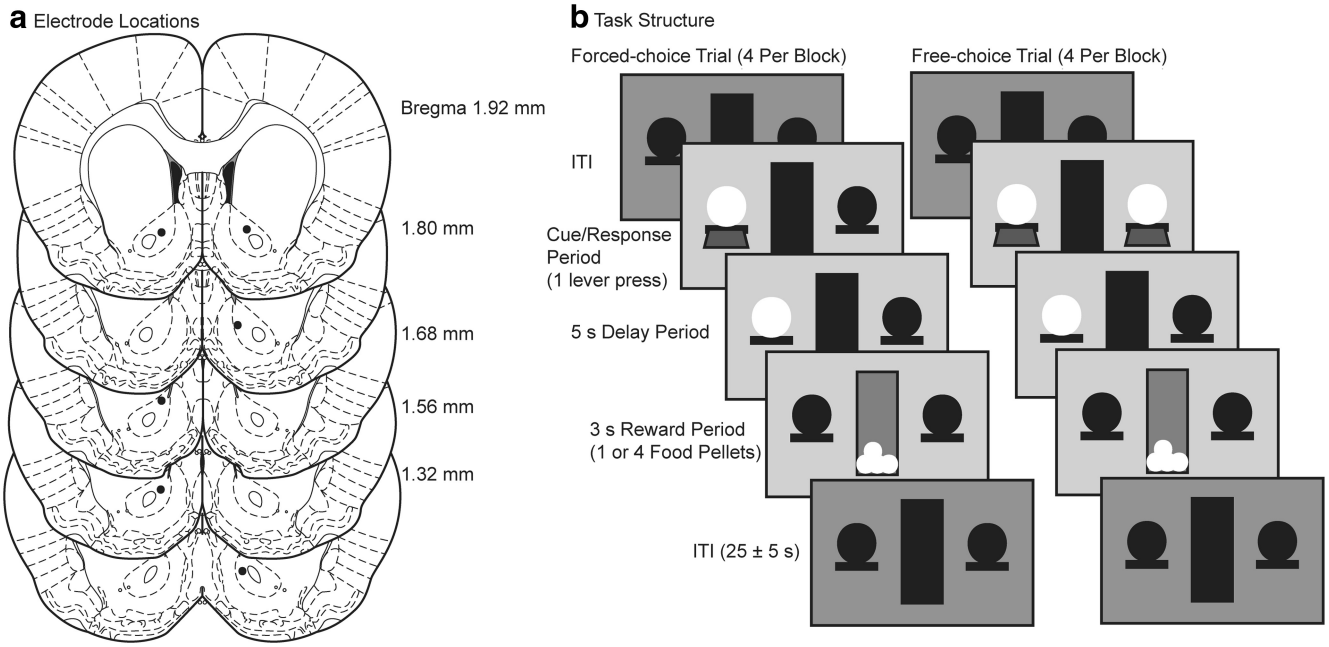

Figure 1. ***a***, Coronal sections showing the locations of chronically implanted electrodes. Brain atlas sections are from Paxinos and Watson (2005). ***b***, The trial structure was the same in both deterministic and probabilistic tasks. A session contained 160 trials in 20 blocks, with four forced-choice (two on each lever) followed by four free-choice trials. In the probabilistic task, lever presses on the 75% lever resulted in four 20 mg food pellets on 75% of trials and one pellet on 25% of trials. Probabilities were reversed for the 25% lever. In the deterministic task, the two levers guaranteed four pellets or one pellet, respectively.

and with Ag/AgCl reference electrodes placed between the skull and meninges. We secured implants to the skull with screws and dental cement built up into a head-cap. Head-caps also included a 6-pin Datamate connector (Harwin) and a nylon head post (Clark et al., 2010). After recovery, we food restricted rats to 85–90% of their postrecovery body weight. Six of the eight rats had electrodes that could detect dopamine, as determined by inspection of cyclic voltammagram (CV) responses to unexpected food pellets in a novel environment. One of these six rats was subsequently excluded due to separation of its head-cap.

Behavioral training. Initial training followed the procedures of Gan et al. (2010) with some modifications. We trained rats to lever press for single 45 mg food pellets (Bioserv) in a free-operant paradigm and then introduced a trial-based task structure. Rats had 30 s to respond after trial start and received a 20 mg food pellet as the reward. Rats then trained on a deterministic task in which pressing one lever always yielded four pellets and the other lever always yielded one pellet. A session included 20 blocks of trials, each containing four forced-choice trials (Fig. 1*b*), during which the rat could only respond on one lever (two trials of each lever per block), and four free-choice trials (Fig. 1*b*), during which the rat could respond on either lever. After rats showed preference for the 4-pellet lever, we introduced a 5 s delay between response and reward. Reward was signaled by tray light onset, followed by feeder activation. The tray light remained illuminated for 3 s before the intertrial interval (ITI) began (25 ± 5 s). When rats preferred the four-pellet lever on at least 41 of 80 free-choice trials and successfully reversed their lever preferences in a subsequent session where the assigned contingencies were reversed, we introduced them to the probabilistic task.

In the probabilistic task rats chose between 75% and 25% lotteries each assigned to one of the levers. The 75% lottery yielded a prize of four pellets with a 75% probability and one pellet otherwise. The 25% lottery yielded a prize of four pellets with a 25% probability and one pellet otherwise. There were four possible lottery-prize combinations with four pellets or one pellet possible from either of the lotteries. We trained rats until they reached a criterion of 10 of the last 12 choices to the more rewarding lever and we performed voltammetry recording on the subsequent session. Recording sessions had an additional criterion that rats choose the more rewarding lever in at least 49 of 80 free-choice trials. If a rat failed to reach criterion on a recording session, it was returned to the 10-of-12-criterion stage. After successful recording sessions, we reversed levers and retrained. This cycle continued until each rat had a minimum of six successful recording sessions with counterbalanced lever assignments.

Four rats completed the probabilistic task, and then we returned them to the deterministic task. Again, the two levers always yielded four pellets or one pellet. To ensure that rats fully expected deterministic rewards, we changed the criterion so that they needed to choose the more rewarding lever on at least 64 of 80 choice trials. After rats reached this more stringent criterion, we performed voltammetry on the subsequent session. We used the same 49-of-80-criterion for voltammetry sessions. The cycle continued until each rat had two voltammetry sessions with counterbalanced lever assignments. One rat required an extra cycle of training and recording because its first voltammetry session had to be discarded due to noise associated with electrical disconnects.

Voltammetry recording and analysis. We recorded dopamine release in the NAc using fast-scan cyclic voltammetry through chronically implanted carbon fiber electrodes (Clark et al., 2010) at a sample rate of 10 Hz. Before voltammetry sessions, we allowed rats extra time in the chamber to condition the electrodes. We delivered a food pellet to the feeder before each session and examined the resultant CV for similarity to a dopamine CV. Behavioral session onsets also elicited dopamine-like responses. We later verified electrodes by calculating the maximum possible correlation coefficient between the presession pellet or session onset recordings and a composite of CVs evoked through electrical stimulation of dopaminergic axons. We included electrodes for which both coefficients were >0.75 and at least one coefficient was >0.80 for all sessions. Six electrodes implanted in the NAc of four rats that completed all behavioral sessions met this criterion (Fig. 1*a*).

We reduced voltammetry data to dopamine oxidation current using background subtraction and principal components regression (PCR) against a training set of electrically evoked dopamine and pH CVs with two principal components (Keithley et al., 2009). Dopamine concentration is proportional to dopamine oxidation current with a factor of ~30 nM/nA (Clark et al., 2010). The background for each trial was the average of the last 10 scans of the delay period. We conducted error analysis on all CVs and excluded CVs for which there was a significant ($\alpha = 0.05$) chance of containing a signal other than dopamine and ΔpH (Keithley et al., 2009). We detrended data from each trial by subtracting a line with its

slope defined by the mean dopamine oxidation currents between −2 to −1 s before reward onset and 10–11 s after reward onset.

Dopamine theories of RPE-based reinforcement learning are primarily motivated by dopaminergic responses to unsignaled rewards (Ljungberg et al., 1992), a finding that has been reproduced with electrochemical measurements (Gan et al., 2010, Clark et al., 2010). Therefore, we isolated the best epoch for testing dopamine for RPE equivalence as the time point likely to contain the maximum response to an unsignaled reward flanked before and after by 0.5 s. We performed PCR as above on presession responses to unsignaled pellets and identified the peak response to unsignaled rewards as the maximum dopamine signal in the 5 s after reward delivery. We then determined the latency for each ($n = 58$) maximum pellet response and counted the number of maximum responses in each 0.1 s bin. The mode of the latency distribution was at 2.0 s, so we used the window 2.0 ± 0.5 s after reward onset in subsequent analyses.

RPE validation. According to the dopamine RPE hypothesis, dopaminergic responses should increase with experienced reward and decrease with predicted reward. Dopamine responses should vary as a function of both the reward received and the reward predicted. Our experimental design allows these two quantities to be modulated independently. We can compute a predicted RPE for each lottery-prize combination for both tasks as the difference between the prize received and the average reward received from a lottery. In the probabilistic task, the average reward received from the 75% lottery is 3.25 pellets and therefore the predicted RPE is 0.75 pellets when four pellets are received and −2.25 when one pellet is received. The average reward received from the 25% lottery is 1.75 pellets and therefore the predicted RPE is 2.25 when four pellets were received and −0.75 when one pellet is received. In the deterministic task, one lever always yielded four pellets and other lever always yielded one pellet. The predicted RPE for both levers is therefore always 0.

We tested the average dopamine signal from the window 2.0 ± 0.5 s after reward onset against an axiomatic RPE model (Caplin and Dean, 2007, Caplin et al., 2010, Rutledge et al., 2010). These three axioms are necessary and sufficient conditions for the entire class of RPE models. The following axioms are the minimum set of assumptions that define any RPE signal δ as a function of a prize z received from a lottery p. $\delta(z)$ is the one-prize "lottery" where prize z is always received.

Axiom 1: Consistent prize ordering:

$$\delta(z, p) > \delta(z', p) \Rightarrow \delta(z, p') > \delta(z', p') \quad (1)$$

Axiom 2: Consistent lottery ordering:

$$\delta(z, p) > \delta(z, p') \Rightarrow \delta(z', p) > \delta(z', p') \quad (2)$$

$$\text{Axiom 3 : No surprise equivalence : } \delta(z') = \delta(z) \quad (3)$$

Conditional logical statements such as Axioms 1 and 2 cannot be proven false if the statement on the left of the conditional arrow is false. Therefore, it is trivially possible for the axioms to be satisfied for a signal that does not distinguish between any lottery or prize conditions, such as noise. Therefore, it is necessary to distinguish between satisfying and strongly satisfying Axioms 1 and 2. By strongly satisfying, we mean that the statement on the left of the conditional arrow is true and the statement on the right of the conditional arrow is also true. This is the experimental condition of interest if one is trying to demonstrate that a neural signal meaningfully meets necessary and sufficient criteria for encoding an RPE signal.

Axiom 1 (consistent prize ordering) tests whether prizes can be consistently ordered by dopamine responses when the expectations (e.g., lotteries) are fixed and the prizes varied. In our experiment, there were two prizes (four pellets and one pellet) and two lotteries (75% and 25% probabilities of four pellets and one pellet otherwise). Axiom 1 is strongly satisfied if four pellets evoked a larger response than one pellet for both the 75% lottery (left side of Axiom 1) and the 25% lottery (right side of Axiom 1). Such a result would imply that receiving four pellets is more rewarding than receiving one pellet, as we would expect. Axiom 1 would be proven false in this experiment if, for example, four pellets evoked a larger response than one pellet for the 75% lottery but one pellet evoked a larger response than four pellets for the 25% lottery. Such a contradiction would violate Axiom 1 and prove that the signal cannot, in principle, represent an RPE signal.

Axiom 2 (consistent lottery ordering) tests whether lotteries can be consistently ordered by dopamine responses when the prizes are fixed and the lotteries varied. Axiom 2 would be strongly satisfied if the 25% lottery produced a larger response than the 75% lottery for both four pellets (left side of Axiom 2) and one pellet (right side of Axiom 2). Such a result would imply that the 75% lottery has a higher predicted reward than the 25% lottery, as we would expect. Axiom 2 would be violated if, for example, the 25% lottery produced a larger response than the 75% lottery to four pellets, but the 75% lottery produced a larger response than the 25% lottery to one pellet. The specific ordering of the prizes (four pellets > one pellet) or lotteries (75% lottery > 25% lottery) is actually irrelevant for the RPE hypothesis as long as ordering is consistent for prizes across lotteries (Axiom 1) and for lotteries across prizes (Axiom 2).

Axiom 3 is falsified if the signal differs between four pellets and one pellet on the deterministic task. For fully anticipated outcomes, the reward prediction error is zero, and therefore the dopamine response should be the same regardless of the prediction. A signal that violates any axiom cannot be an RPE representation. A signal for which all three axioms are true necessarily encodes an RPE-equivalent representation.

We first tested the axioms by counting the number of observations for which prize and lottery ordering were consistent and thus satisfied the axioms. We then calculated the probability that observed counts were due to chance using the cumulative binomial distribution for a probability of 0.5, with 12 (6 electrodes by 2 lotteries or 6 electrodes by 2 prizes) observations. We conducted a similar analysis on Axiom 3 violations for which the signal to 4 fully predicted pellets was greater than the signal to 1 fully predicted pellet and calculated the probability that observed counts were due to chance using the cumulative binomial distribution for a probability of 0.5 with 6 observations.

To more strictly test the signals against the axiomatic model, we conducted four planned paired t tests on dopamine release in the probabilistic task. Significant differences in opposite directions for four pellets versus one pellet for the 75% lottery and for four pellets versus one pellet for the 25% lottery would indicate an Axiom 1 violation. Significant differences in opposite directions for the 75% versus 25% lottery for four pellets and for the 75% versus 25% lottery for one pellet would indicate an Axiom 2 violation. Significant differences in the same direction for either pair of comparisons indicate that the axiom is strongly satisfied. We corrected α for multiple comparisons using the Holm–Bonferroni procedure. We used a paired t test for the effect of prize on mean dopamine release in the deterministic task for the same epoch. A significant effect of prize in the deterministic task would indicate an Axiom 3 violation.

We also fit the relationship between predicted RPE and dopamine signal using a line with a single slope and using a piecewise linear function with a single intercept but separate slopes for the positive and negative domains and we used an F test to test for a significant improvement in fit.

Epoch analysis. We conducted a set of *post hoc* analyses to determine the effect of window size and window center on the RPE equivalency of the signal and the linear relationship between model RPE and dopamine concentrations. We performed the same counts of strongly satisfying observations and sets of 4 t tests for Axioms 1 and 2 on all possible 0.5, 1.1, and 1.9 s time windows (10 samples/s) between 0.1 and 5 s after reward onset. For Axioms 1 and 2, we calculated the conjunction p-value from the four t tests as largest of the four p-values taken to the fourth power. This p-value provides a conservative estimate of the probability of observing the expected combination of prize and lottery effects given the global null hypothesis that the true signal is not sensitive to either prize or lottery effects (Rutledge et al., 2010). We corrected all conjunction p-values for multiple comparisons by multiplying by the number of t tests (four) and the number of time windows analyzed. We used MATLAB (MathWorks) for all data processing and statistics. $\alpha = 0.05$ for all tests except where corrected for multiple comparisons.

Histology. We anesthetized rats with 150 mg/kg ketamine. We performed electrolytic lesions through the electrodes and perfused the rats through the heart with saline, followed by paraformaldehyde (PFA; 40 g/L) in PBS. We removed brains and stored them for at least 24 h in 40 g/L

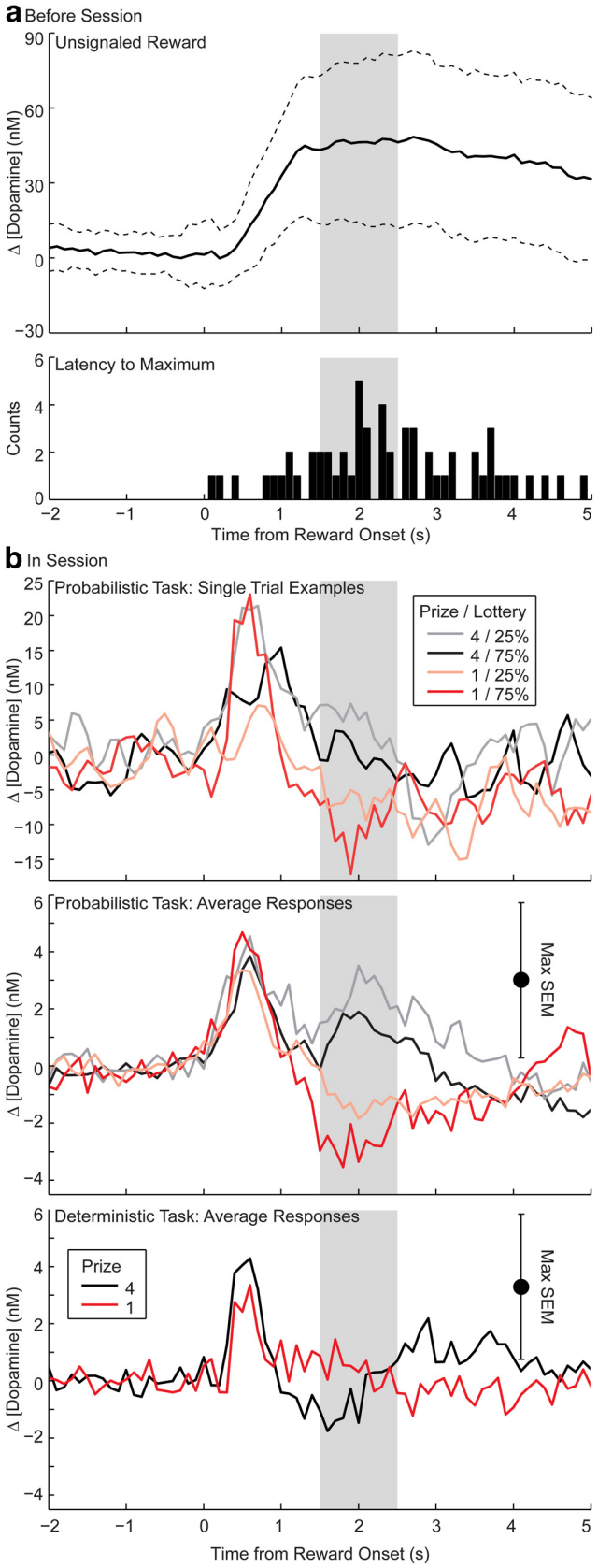

Figure 2. ***a***, Top, Mean ± SD dopamine response to an unsignaled food pellet reward delivered before the beginning of a behavioral session ($n = 58$, 6 electrodes). Bottom, Latency to maximum dopamine signal for each unsignaled reward presentation. ***b***, Top, Reward-evoked changes in dopamine concentration recorded at one electrode within a single behavioral session for the four lottery-outcome combinations in the probabilistic task. For the purpose of illustration, traces were smoothed with 3-point running average. Middle and bottom, Average reward-evoked dopamine concentration for $n = 6$ electrodes for each of the lottery-outcome combinations on forced trials during the probabilistic (upper) and deterministic (lower) tasks. Shading indicates the epoch used in subsequent analyses.

PFA in PBS at 4°C. We saturated brains in 300 g/L sucrose solution at 4°C and froze them on dry ice. We cut frozen brains in 50 μm sections on a cryostat and mounted sections on microscope slides. We then Nissl stained mounted sections and located lesion sites using an adult rat brain atlas (Paxinos and Watson, 2005).

Results

Behavior

Behavior on choice trials revealed preferences for the lever that returned the higher average reward. During sessions of the probabilistic task used for voltammetry analysis, rats preferred the 75% lottery, which yielded four pellets with a 75% probability and one pellet otherwise. Rats chose the 75% lottery on 71.93 ± 1.09% (mean ± SD) of choice trials. Later, on the simpler deterministic version of the task in which the levers deterministically yielded four pellets or one pellet, respectively, preferences for the more rewarding lever were enhanced. Rats chose the four-pellet lever on 85.45 ± 5.93% of choice trials, indicating a high degree of preference.

Epoch selection

Dopamine concentration in the NAc (Fig. 1*a*) rapidly rose in response to unsignaled rewards delivered before the beginning of behavioral sessions (Fig. 2*a*). The mode of the distribution of latencies to maximum pellet-evoked dopamine was 2.0 s. Therefore, we performed subsequent analyses on mean dopamine responses from 2.0 ± 0.5 s after reward onset.

Test of axiomatic RPE model

Phasic dopamine release 2.0 ± 0.5 s after reward onset in the NAc met the three necessary and sufficient conditions for RPE equivalence from the axiomatic RPE model (Caplin and Dean, 2007). The average dopamine concentration across electrodes ($n = 6$ electrodes) during the reward-delivery period was characterized by a positive change in dopamine to four pellets and a negative change to one pellet for either lottery. Dopamine responses were also ordered with respect to lottery, with dopamine responses to outcomes from 25% lottery greater than from 75% lottery for both prizes (Fig. 2*b*).

Mean dopamine responses from 2.0 ± 0.5 s after reward onset in the probabilistic task were consistent with Axiom 1. For Axiom 1 to be satisfied, the same prize must evoke the larger dopamine response when it is the outcome from either lottery. For Axiom 1 to be violated, prize ordering must differ between lotteries. Figure 3*a* shows that for all 12 observations (6 electrodes by 2 lotteries), the 4-pellet prize evoked a larger signal than the 1-pellet prize, as all points are above the equivalence line, and this effect was significant ($p < 0.001$, binomial test). Therefore, the ordering of dopamine responses from all electrodes was consistent with Axiom 1.

Dopamine responses from the probabilistic task were also consistent with Axiom 2. For Axiom 2 to be satisfied, the same lottery must produce the larger response for both prizes. For Axiom 2 to be violated, the ordering of responses to the lotteries must differ between prizes. Figure 3*b* shows that for 10 of 12 observations, the 25% lottery produced a larger signal than the 75% lottery, whereas the 75% lottery produced a larger signal on only 2 of 12 observations. Overall, this difference was significant ($p = 0.0193$, binomial test). Therefore, the ordering of observed dopamine responses was consistent with Axiom 2.

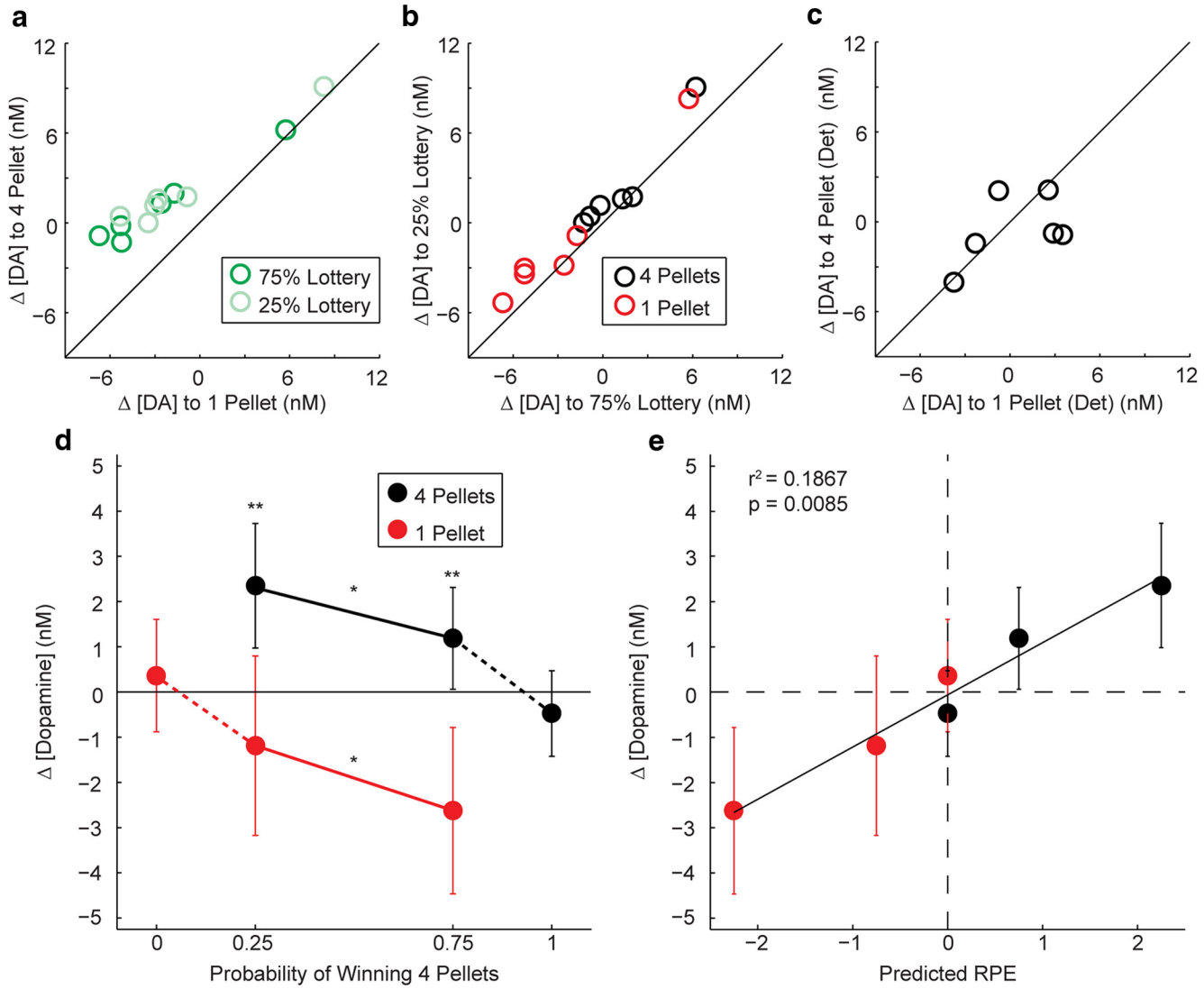

Figure 3. ***a***, Mean dopamine release to four pellets is plotted against mean dopamine release to one pellet after both 75% and 25% lottery forced-choice trials for each electrode. Points above the line indicate greater dopamine release to four pellets than to one pellet. ***b***, Mean dopamine release on 25% lottery trials is plotted against mean dopamine release on 75% lottery trials for both prizes for each electrode. Points above the line indicate greater dopamine concentrations when pellets are received from the 25% lottery than from the 75% lottery. ***c***, Mean dopamine release to four pellets is plotted against mean dopamine to one pellet on the deterministic task for each electrode. Responses are heterogeneously distributed around the equivalence line. ***d***, The mean dopamine release from ***a***–***c*** for $n = 6$ electrodes are shown for both the probabilistic and deterministic tasks. Ordering of signals satisfies the axiomatic RPE model. ***e***, Mean ± SEM dopamine release for each possible lottery-prize combination across the two tasks is plotted against the predicted RPE, calculated as the difference between reward magnitude and average lottery outcome. The line indicates a significant linear relationship. *$p < 0.05$ for comparison between lotteries; **$p < 0.01$ for comparison between prizes, paired t test.

Dopamine responses in the deterministic task showed no systematic deviation. For Axiom 3 to be satisfied, responses must not differ between prizes. In Figure 3*c*, responses are heterogeneously distributed around the equivalence line with four values above the line and two values below the line. Therefore, the observed deviations were not different from those expected due to chance ($p = 0.34$, binomial test).

Additional analysis confirmed that dopamine concentration satisfied the axiomatic RPE model (Fig. 3*d*). Four planned paired t tests with α levels corrected using the Holm–Bonferroni procedure revealed results for Axioms 1 and 2 consistent with the scatter plots. The dopamine responses were larger for four pellets than for one pellet for both the 75% lottery ($t_5 = 5.0532$, $p = 0.0039$) and the 25% lottery ($t_5 = 5.0824$, $p = 0.0038$), satisfying Axiom 1. The dopamine responses for 25% lottery trials were significantly larger than responses for 75% lottery trials for both the 4-pellet prize ($t_5 = 2.6796$, $p = 0.044$) and the 1-pellet prize ($t_5 = 3.4699$, $p = 0.018$), satisfying Axiom 2. Finally, a single paired t test of responses from the deterministic task revealed that there was not a significant effect of prize when rewards were fully predicted ($t_5 = 0.7576$, $p = 0.48$), satisfying Axiom 3.

Test for asymmetry of RPE signals

Curve fits of the relationship between predicted RPEs calculated as the difference between reward size and average outcome from each lottery and measured dopaminergic responses revealed no asymmetry in encoding of positive and negative RPEs. Least-squares regression showed that a line fit the data with a y-intercept of -0.0022 ± 0.0188 and a slope of 0.0384 ± 0.0138 ($r^2 = 0.1867$, $F_{(1,34)} = 7.8069$, $p = 0.0085$; Fig. 3*e*). A piecewise linear function with a single intercept but separate slopes for the positive and negative domains did not improve the fit ($F_{(1,33)} = 0.0049$, $p = 0.94$) and the slopes were nearly identical for the positive and negative domains

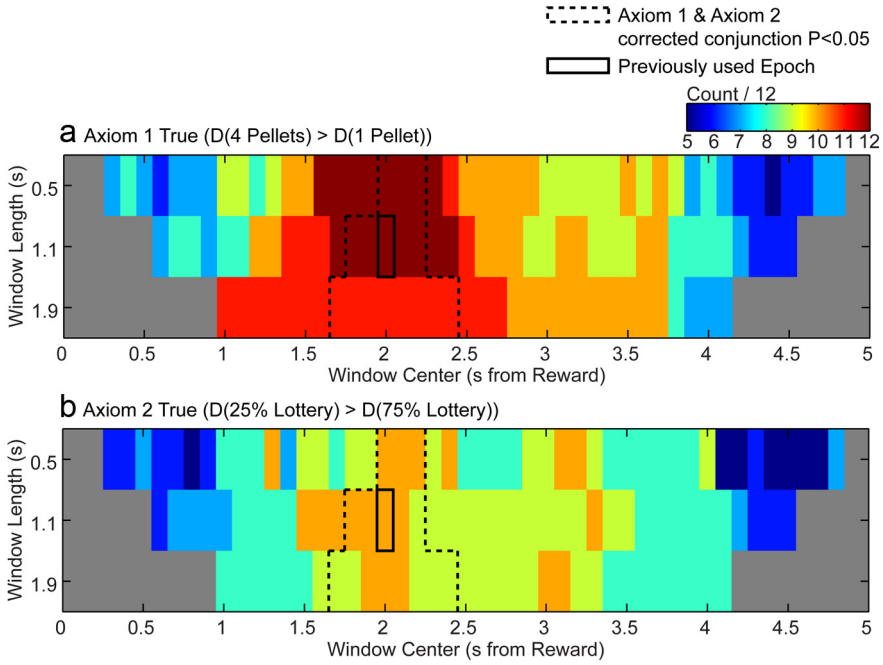

Figure 4. Epoch analysis for all 118 possible 0.5, 1.1, and 1.9 s windows between 0.1 and 5 s after reward onset. ***a***, Colors indicate the number of lottery/electrode combinations for which the dopamine signal to four pellets was greater than the dopamine signal for one pellet for each time window. Counts >9 are consistent with Axiom 1. ***b***, Colors indicate the number of prize/electrode combinations for which the dopamine signal to 25% lottery outcomes was greater than the dopamine signal to 75% lottery outcomes. Counts >9 are consistent with Axiom 2. Dashed lines indicate the set of time windows for which the corrected conjunction *p*-value for *t* tests of Axioms 1 and 2 is <0.05. The solid line indicates the time window analyzed in Figure 3.

(RPE < 0: slope = 0.0399 ± 0.0248, RPE > 0: slope = 0.0370 ± 0.0248). As for the line with a single slope, the *y*-intercept was near 0 (−0.0008 ± 0.0280).

Window analysis

Although our initial analysis was on the epoch determined by the peak response to unsignaled rewards outside the task, we subsequently conducted an unbiased test of the entire period between reward delivery and the start of the ITI to test whether the RPE encoding generalized to other times. *Post hoc* analysis of all 118 possible 0.5, 1.1, and 1.9 s windows (10 samples/s) between 0 and 5 s after reward revealed that there was a narrow range of windows for which the ordering of dopamine responses was consistent with the axiomatic RPE model and that window center was more important than window length in determining whether the mean dopamine response could be an RPE. A broad range of time windows had 10 or more electrode/lottery combinations for which there was a greater dopamine response to four pellets than to one pellet for a given lottery and electrode (Fig. 4*a*), which is consistent with Axiom 1. In contrast, a smaller set of time windows centered around 2 s after reward delivery had at least 10 electrode/prize combinations for which there was a greater dopamine response to the 25% lottery than the 75% lottery (Fig. 4*b*), consistent with Axiom 2. Corrected conjunction *p*-values of paired *t* tests for Axioms 1 and 2 also demonstrate that windows of all 3 durations centered around 2 s after reward delivery have dopamine signals consistent with Axioms 1 and 2.

Discussion

In the present study, we used operant decision-making tasks and the axiomatic model of Caplin and Dean (2007) to identify an RPE represented by phasic dopamine release in the NAc. We designed the task so that there were two lotteries that each provided either of two rewards probabilistically. Because reward variance did not differ between the two lotteries and because rewards were delivered on every trial, this design necessarily avoids confounding effects of salience associated with uncertainty in the form of reward variance (Fiorillo et al., 2003) or reward omissions (Esber and Haselgrove, 2011). Both lotteries produced either one or four food pellets with a probability of either 0.75 or 0.25. The reward probability assigned to prizes was reversed between lotteries. Strongly satisfying the axiomatic model, NAc dopamine signals were coherently modulated by prize and by lottery in the probabilistic task, whereas prizes did not significantly modulate dopamine signals in the deterministic task. By strongly satisfying all three conditions of the axiomatic model (Caplin and Dean, 2007), phasic dopamine release in the NAc meets the necessary and sufficient conditions for RPE equivalence. This observation does not necessitate that the RPE signal is the only signal carried by dopamine and it is important to note that our analysis focused on a particular epoch. We detrended the data to remove signaling on longer timescales than this epoch, appropriate to the design of the task. However, it has been shown that meaningful information is encoded in extracellular dopamine concentrations in both the rodent (Howe et al., 2013) and human brain (Kishida et al., 2011). *Post hoc* analysis of the set of 118 possible 0.5, 1.1, and 1.9 epochs centered between 0 and 5 s after reward delivery showed that this RPE signal is specific to windows centered at 2.0 or 2.1 s after reward onset, but that epoch length did not influence the result.

It should be noted that the latency of the dopaminergic RPE signal is much greater than that shown in electrophysiological studies on putative midbrain dopamine neurons (Montague et al., 1996, Schultz et al., 1997, Waelti et al., 2001), but it is consistent with the latencies of reward-evoked and cue-evoked changes in extracellular dopamine concentration measured in other studies (Clark et al., 2010, Flagel et al., 2011). The difference in latencies between electrophysiological studies and voltammetry studies are primarily attributable to diffusion of dopamine from release sites to the electrode surface. Because dopamine acts as a volume transmitter at extrasynaptic receptors, diffusion is expected to slow the temporal dynamics of dopamine signal transduction in regions such as the nucleus accumbens (Venton et al., 2003).

Outside the epoch centered around 2 s after reward onset, there was one other component to the dopamine signal that was inconsistent with an RPE. The peak observed in the first second after reward onset was temporally resolved from the RPE signal and was not modulated by reward expectation, making it incompatible with RPE coding. The short latency and the lack of modulation between conditions suggest that it is related to a task feature that does not differ between reward and lottery conditions or between the probabilistic and deterministic tasks. Such

features include the extinguishing of the cue light at the end of the 5 s delay period, the illumination of the feeder light, or the click of the feeder for the first food pellet. These three features might form a compound stimulus that signals the end of the delay period and the beginning of the reward phase of the trial. A persistent dopamine response to this stimulus might be present even in well trained rats if they are unable to accurately time out a 5 s delay, causing the onset of the reward period to be a surprising event. However, because this signal did not change with reward expectation, it is more likely to encode an attentional process rather than a pure economic process.

Our data also allowed us to examine a second major issue in the study of dopamine and learning: the nature of the relationship between dopamine release and RPEs across the positive and negative domains. Bayer and Glimcher (2005) reported that the firing rates of putative dopamine neurons in the SNc correlated with positive but not negative RPEs. They hypothesized that a second system might exist that encoded these negative RPEs. Subsequent analyses (Bayer et al., 2007) revealed that the magnitude of negative RPEs was correlated with the duration of the extended interspike pause that follows reward delivery. Long pauses provided a signal from which negative, but not positive, RPEs could be extracted. This analysis is of particular relevance because extracellular dopamine concentration, through volume transmission, might reflect temporally integrated spike bursts and pauses due to the interplay among release, diffusion, and uptake. This interplay can exist in the case of volume transmission because clearance is slower than in synaptic transmission (Garris et al., 1994). Mathematical models emphasize the nonlinearity in the relationship between firing rate of dopamine neurons and extracellular dopamine concentration conferred by uptake (Wightman et al., 1988), diffusion (Venton et al., 2003) and short-term elastic modulation of release probability (Montague et al., 2004), as well as the impact of dopamine release and uptake on dopamine receptor occupancy (Dreyer et al., 2010). Therefore, we tested the relationship between dopamine signals and predicted RPEs calculated as the difference between the prize magnitude and average outcome on the lottery. We found that the relationship between RPEs and extracellular dopamine concentration was best fit by a straight line with a *y*-intercept near zero. Moreover, allowing slopes to vary for positive and negative RPEs in a piecewise linear function did not significantly improve the fit. These findings show that positive and negative dopaminergic RPEs were represented symmetrically at the level of dopamine concentration without the rectification observed in firing rate of dopamine neurons (Bayer and Glimcher, 2005). Although it does not rule out the possibility that opponent neural systems cooperate to compute motivationally relevant variables, the present work demonstrates that dopamine transmission is sufficient to provide a bidirectional teaching signal to the NAc and that a complementary opponent system to deliver negative RPEs would not be necessary in the range of RPEs we measured.

References

Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47:129–141. CrossRef Medline

Bayer HM, Lau B, Glimcher PW (2007) Statistics of midbrain dopamine neuron spike trains in the awake primate. J Neurophysiol 98:1428–1439. CrossRef Medline

Caplin A, Dean M (2007) Dopamine, reward prediction error, and economics. Quarterly Journal of Economics 123:663–701.

Caplin A, Dean M, Glimcher PW, Rutledge RB (2010) Measuring beliefs and rewards: a neuroeconomic approach. Quarterly Journal of Economics 125:923–960. CrossRef

Clark JJ, Sandberg SG, Wanat MJ, Gan JO, Horne EA, Hart AS, Akers CA, Parker JG, Willuhn I, Martinez V, Evans SB, Stella N, Phillips PEM (2010) Chronic microsensors for longitudinal, subsecond dopamine detection in behaving animals. Nat Methods 7:126–129. CrossRef Medline

Dreyer JK, Herrik KF, Berg RW, Hounsgaard JD (2010) Influence of phasic and tonic dopamine release on receptor activation. J Neurosci 30:14273–14283. CrossRef Medline

Esber GR, Haselgrove M (2011) Reconciling the influence of predictiveness and uncertainty on stimulus salience: a model of attention in associative learning. Proc Biol Sci 278:2553–2561. CrossRef Medline

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

Flagel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, Willuhn I, Akers CA, Clinton SM, Phillips PEM, Akil H (2011) A selective role for dopamine in stimulus-reward learning. Nature 469:53–57. CrossRef Medline

Gan JO, Walton ME, Phillips PEM (2010) Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. Nat Neurosci 13:25–27. CrossRef Medline

Garris PA, Ciolkowski EL, Pastore P, Wightman RM (1994) Efflux of dopamine from the synaptic cleft in the nucleus accumbens of the rat brain. J Neurosci 14:6084–6093. Medline

Glimcher PW (2011) Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proc Natl Acad Sci U S A 108:15647–15654. CrossRef Medline

Howe MW, Tierney PL, Sandberg SG, Phillips PEM, Graybiel AM (2013) Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. Nature 500:575–579. CrossRef Medline

Keithley RB, Heien ML, Wightman RM (2009) Multivariate concentration determination using principal component regression with residual analysis. Trends Analyt Chem 28:1127–1136. CrossRef Medline

Kishida KT, Sandberg SG, Lohrenz T, Comair YG, Sáez I, Phillips PEM, Montague PR (2011) Sub-second dopamine detection in human striatum. PLoS One 6:e23291. CrossRef Medline

Ljungberg T, Apicella P, Schultz W (1992) Responses of monkey dopamine neurons during learning of behavioral reactions. J Neurophysiol 67:145–163. Medline

Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16:1936–1947. Medline

Montague PR, McClure SM, Baldwin PR, Phillips PE, Budygin EA, Stuber GD, Kilpatrick MR, Wightman RM (2004) Dynamic gain control of dopamine delivery in freely moving animals. J Neurosci 24:1754–1759. CrossRef Medline

Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H (2006) Midbrain dopamine neurons encode decisions for future action. Nat Neurosci 9:1057–1063. CrossRef Medline

Paxinos G, Watson C (2005) The rat brain in stereotaxic coordinates, Ed 5. London: Elsevier Academic.

Phillips PEM, Walton ME, Jhou TC (2007) Calculating utility: Preclinical evidence for cost-benefit analysis by mesolimbic dopamine. Pyschopharmacology (Berl) 191:483–495. CrossRef Medline

Rutledge RB, Dean M, Caplin A, Glimcher PW (2010) Testing the reward prediction error hypothesis with an axiomatic model. J Neurosci 30:13525–13536. CrossRef Medline

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599. CrossRef Medline

Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH (2013) A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci 16:966–973 CrossRef Medline

Sutton RS (1988) Learning to predict by the methods of temporal differences. Machine Learning 3:9–44.

Venton BJ, Zhang H, Garris PA, Phillips PEM, Sulzer D, Wightman RM (2003) Real-time decoding of dopamine concentration changes in the caudate-putamen during tonic and phasic firing. J Neurochem 87:1284–1295. CrossRef Medline

Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. Nature 412:43–48. CrossRef Medline

Wightman RM, Amatore C, Engstrom RC, Hale PD, Kristensen EW, Kuhr WG, May LJ (1988) Real-time characterization of dopamine overflow and uptake in the rat striatum. Neuroscience 25:513–523. CrossRef Medline