

28. We thank T. Pekar, J. Park, and J. Guo for technical assistance. Supported by Canadian Institutes of Health Research (MOP74650, MOP77561) and EHLB Foundation grants to S.A.J. and P.W.F., National Alliance for Research on Schizophrenia and Depression fellowship to J.H.H., travel fellowship (France-Canada Research Foundation), Agence Nationale de la Recherche (France)

grant (06-NEURO-027-03) to B.B., and Restracom Fellowship from the Hospital for Sick Children to A.P.Y. and J.H.H.

Supporting Online Material
www.sciencemag.org/cgi/content/full/323/5920/1492/DC1
Materials and Methods

SOM Text
Figs. S1 to S9
References
Movie S1

4 August 2008; accepted 15 January 2009
10.1126/science.1164139

Human Substantia Nigra Neurons Encode Unexpected Financial Rewards

Kareem A. Zaghoul,^{1*} Justin A. Blanco,² Christoph T. Weidemann,³ Kathryn McGill,¹ Jurg L. Jaggi,¹ Gordon H. Baltuch,¹ Michael J. Kahana^{3*}

The brain's sensitivity to unexpected outcomes plays a fundamental role in an organism's ability to adapt and learn new behaviors. Emerging research suggests that midbrain dopaminergic neurons encode these unexpected outcomes. We used microelectrode recordings during deep brain stimulation surgery to study neuronal activity in the human substantia nigra (SN) while patients with Parkinson's disease engaged in a probabilistic learning task motivated by virtual financial rewards. Based on a model of the participants' expected reward, we divided trial outcomes into expected and unexpected gains and losses. SN neurons exhibited significantly higher firing rates after unexpected gains than unexpected losses. No such differences were observed after expected gains and losses. This result provides critical support for the hypothesized role of the SN in human reinforcement learning.

Theories of conditioning and reinforcement learning postulate that unexpected rewards play an important role in allowing an organism to adapt and learn new behaviors (1, 2). Research on nonhuman primates suggests that midbrain dopaminergic neurons projecting from the ventral tegmental area and the pars compacta region of the SN encode unexpected reward signals that drive learning (3–6). These dopaminergic neurons are phasi-

cally activated in response to unexpected rewards and depressed after the unexpected omission of reward (7–9), and they are major inputs to a larger basal ganglia circuit that has been implicated in reinforcement learning across species (10–15).

The response of these neurons to rewards has not been directly measured in humans. We recorded neuronal activity in human SN while patients undergoing deep brain stimulation (DBS)

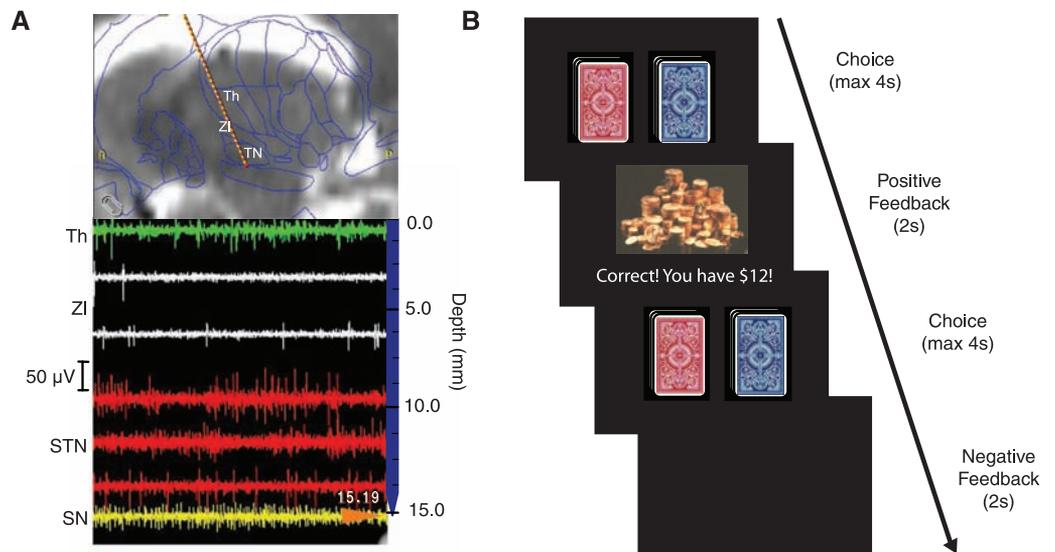
surgery for Parkinson's disease performed a probability learning task. Patients with Parkinson's disease show impaired learning from positive and negative feedback in cognitive tasks (16–18), probably because of the degenerative nature of their disease and the decreased number of dopaminergic neurons capable of mounting phasic changes in activity in response to reward signals (17–19). We sought to capture remaining viable dopaminergic SN cells in our patients and determine whether they exhibit responses modulated by reward expectation.

We used microelectrode recordings to measure intraoperative activity of SN in 10 Parkinson's patients (6 men, 4 women, mean age of 61 years) undergoing DBS surgery of the subthalamic nucleus (STN) while they engaged in a probability learning task. We rewarded participants in the task with virtual financial gains to motivate learning. We identified SN by anatomic location and its unique firing pattern (Fig. 1A) (20). The learning task involved choosing between a red and a blue deck of cards presented on a computer screen (Fig. 1B). We informed partici-

¹Department of Neurosurgery, University of Pennsylvania, Philadelphia, PA 19104, USA. ²Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA. ³Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA.

*To whom correspondence should be addressed. E-mail: zaghoul@uphs.upenn.edu (K.A.Z.); kahana@psych.upenn.edu (M.J.K.)

Fig. 1. (A) Intraoperative plan for DBS surgery with targeting of the STN. Microelectrodes are advanced along a tract through the anterior thalamic nuclei (Th), zona incerta (ZI), STN, and into the SN to record neural activity. Each anatomical region is identified by surgical navigation maps overlaid with a standard brain atlas (top) and by its unique firing pattern and microelectrode position (bottom). Depth measurements on the right of the screen begin 15 mm above the pre-operatively identified target, the inferior border of STN. In this example, the microelectrode tip lays 0.19 mm below the target. A, anterior; P, posterior. **(B)** Probability learning task. Participants are presented with two decks of cards on a computer screen. They are instructed to repeatedly draw cards from either deck to determine which deck yields the higher reward probability. Participants are given up to four seconds for each draw. After each draw, positive or negative feedback is presented for two seconds. Decks are then immediately presented on the screen for the next choice.



pants that one of the two decks carried a higher probability of yielding a financial reward than the other. Participants were instructed to repeatedly draw cards from either deck to determine which yields a higher return (high reward rate 65%, low reward rate 35%) (20). If the draw of a card yielded a reward, a stack of gold coins was displayed along with an audible ring of a cash register and a counter showing accumulated virtual earnings. If the draw did not yield a reward or if no choice was made, the screen turned blank and participants heard a buzz. Participants completed 91.5 ± 13.3 (mean \pm SD) trials during the 5-min experiment.

We examined learning rates for the experiment (Fig. 2A) (20). Once a participant learns which deck has the higher payoff probability, he or she should preferentially choose that deck. On average, the rate with which participants chose the higher-probability deck improved from $52.5 \pm$

4.9% (mean \pm SEM) to $70.0 \pm 4.4\%$ over the course of the experiment.

We sought to determine when observed feedback differed from expected feedback. In previous quantitative models of retention, memory performance falls off approximately as a power function of the retention interval, decaying rapidly in the short term and slowly in the long term (21, 22). Such a functional relation weighs recent experiences more heavily in determining the expected probability of a reward. Here, we use a power function to define the expected reward from a given deck as a function of reward history. Choosing a particular deck, d , on the n th trial will yield an expected reward, $E_d[n]$, defined as

$$E_d[n] = 0.5 + 0.5 \sum_{i=1}^{n-1} R_d[i] \alpha i^{-\tau} \quad n = 2, \dots, N \quad (1)$$

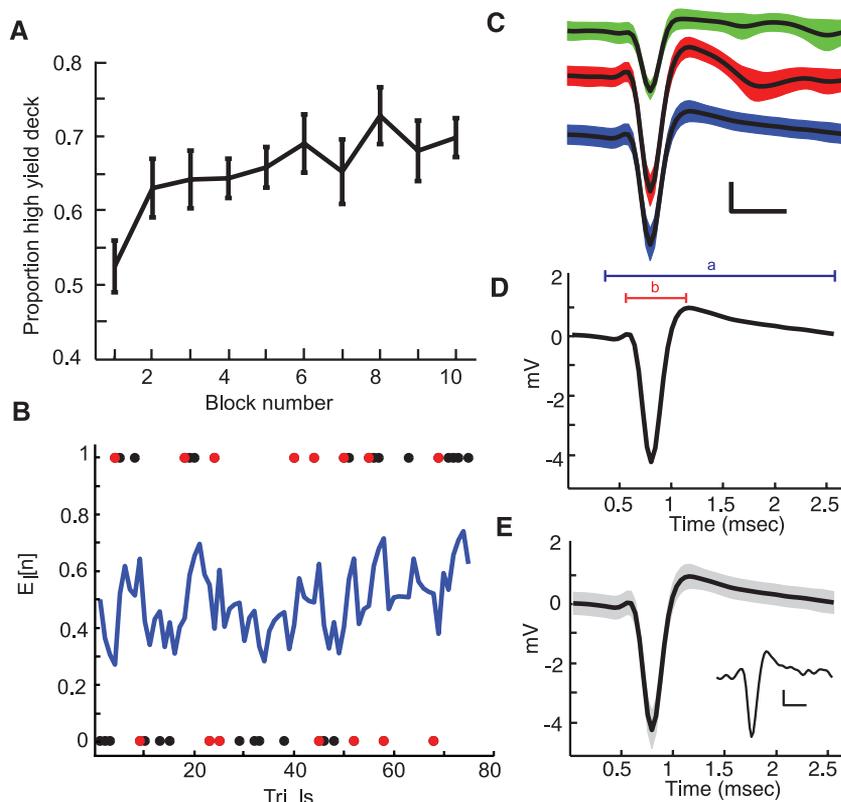


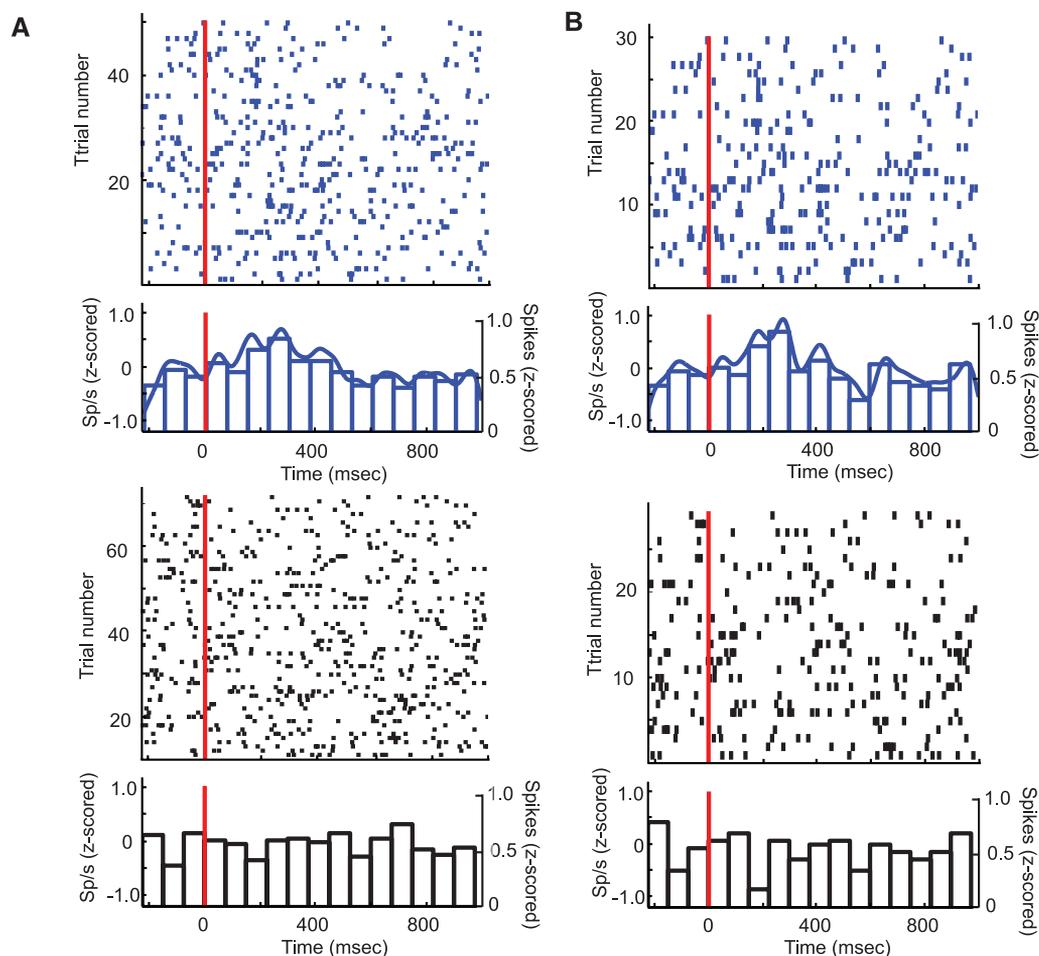
Fig. 2. (A) Learning rates are quantified by dividing the total number of trials (draws from the decks) into 10 equally sized blocks and determining how often participants correctly chose the (objectively) better deck during that block. Trace represents mean learning rate across all participants. Error bars represent SEM. (B) Expected reward associated with one deck in a single experiment. For each trial, we show the expected reward computed for the left deck, $E_l[n]$ (blue line) (Eq. 1). The outcome of each trial when this deck was selected is shown as a circle. Circles having value 1 represent positive outcomes, whereas circles having value 0 represent negative outcomes. Black circles denote expected outcomes, and red circles denote unexpected outcomes. We base our analysis on unexpected outcomes. (C) Mean waveforms of three unique spike clusters from one participant are shown in black, with SD colored for each cluster. Scale bar, 10 mV and 0.5 msec. (D) For each identified cluster, we calculated the average time from the beginning of the spike waveform to its return to baseline (a) and the average time between the two positive peaks of the waveform (b). We restricted our analysis to those clusters that had average baseline widths greater than 2 msec and peak-to-peak widths greater than 0.8 msec. (E) Mean ($n = 4703$) waveform of spikes from a single cell from one participant is shown in black with SD in gray. (Inset) Example waveform. Inset scale bar, 1 mV and 1 msec.

for that deck. $R_d[n]$ is defined as the feedback of the n th trial (total number of trials = N) for deck d and has a value of 1 for positive feedback, -1 for negative feedback, and 0 for trials when deck d was not selected. Expectation is thus computed as a weighted sum of previous outcomes, where the weights fall off with the power function determined by τ . We set α such that the weights of the power function approximate one over infinite trials for a given τ . This ensures an unbiased estimate of the effect of prior outcomes on expectation and limits expectation to the range between zero and one. We fit Eq. 1 to the sequence of choices and rewards observed in each experimental session to determine the best-fitting τ for every participant [$\tau = 1.68 \pm 0.32$ (mean \pm SEM)] (20). Based on the best fitting τ values in this model, participants selected the deck with the higher expected reward on $74.9 \pm 3.1\%$ of the trials. We used this model of expected reward to classify the feedback associated with each trial into one of four categories: (i) unexpected gains, (ii) unexpected losses, (iii) expected gains, and (iv) expected losses (23). The expected reward associated with one deck for a single experiment from a single participant is shown in Fig. 2B as a function of trial number n .

We extracted and sorted single-unit activity captured from SN microelectrode recordings to find 67 uniquely identified spike clusters (3.94 ± 0.6 clusters per recording) (Fig. 2C). To restrict our analysis to dopaminergic cells, we applied to each cluster stringent criteria pertaining to firing rate, spike morphology, and response to feedback, derived from previous studies (Fig. 2D) (20, 24, 25). Ultimately, we retained 15 putatively dopaminergic spike clusters, hereafter referred to as cells, for analysis [0.88 ± 0.21 (mean \pm SEM) cells and $21.4 \pm 6.5\%$ of total spikes per recording; 10 microelectrode recordings contributed to this subset]. Average recorded waveforms from one cell and an example of an individual waveform are shown in Fig. 2E.

Representative spike activity recorded from a single SN cell in a single participant is shown in Fig. 3. We quantified the differences in spike activity during 225-msec non-overlapping intervals (20), focusing our analyses on the interval between 150 and 375 msec after the onset of feedback. Preliminary analyses demonstrated that this interval was particularly responsive (20), and this interval is consistent with response latencies shown in animal studies (6, 7). Raw spike count increased in response to positive feedback and decreased in response to negative feedback during this interval [$F_{1,110} = 4.6$, mean squared error (MSE) = 1.1, $P = 0.04$] (Fig. 3A). Fig. 3B shows spike activity during trials associated with unexpected gains and losses, recorded from the same SN cell. The difference in activity between responses to unexpected gains and losses during this interval was statistically significant [$F_{1,57} = 6.9$, MSE = 1.8, $P = 0.01$] and notably clearer than the difference between positive and negative feedback.

Fig. 3. (A) Spike raster for a single experiment from one participant. Individual spike activity recorded from SN for trials during positive (blue) and negative (black) feedback is shown for each trial as a function of time. Below each spike raster is the average z-scored continuous-time firing rate (continuous trace) and histogram (bars, 75-msec intervals). The red vertical line indicates feedback onset. **(B)** Individual spike activity, recorded from the same cell as shown in Fig. 3A, for trials in response to unexpected gains (blue) and losses (black) is shown for each trial as a function of time.



To determine how SN neurons encode behavioral feedback across participants, we pooled results for all cells meeting our inclusion criteria. We compared continuous-time firing rates and spike histograms for each SN cell to its baseline spiking activity to generate average z-scored continuous-time firing rates and histograms for each cell (Fig. 3) (20). We compared neural responses to unexpected and expected positive and negative feedback using a three-way analysis of variance for the interval between 150 and 375 msec after feedback onset (20). We found a significant difference between responses to positive and negative feedback in z-scored firing rate [$F_{1,14} = 9.3$, MSE = 29, $P = 0.0082$] and spike counts [$F_{1,14} = 16$, MSE = 16, $P < 0.005$]. In addition, we found that this main effect of feedback was modulated by a significant interaction with expectation [$F_{1,14} = 11.3$, MSE = 26, $P < 0.005$] for continuous-time firing rate and [$F_{1,14} = 15.0$, MSE = 17, $P < 0.005$] for spike count. The other post-feedback intervals (20) exhibited no significant differences between responses to positive and negative feedback. In addition, we found no significant change in activity in response to deck presentation itself [supporting online material (SOM) text].

To further investigate the strong modulatory effect of expectation, we examined the pooled

activity across participants in response to unexpected gains and losses only. During the same post-feedback interval (gray shaded region, Fig. 4A) (20), spike rates in response to unexpected gains were significantly greater than spike rates in response to unexpected losses [$F_{1,14} = 16.5$, MSE = 49, $P < 0.001$] (Fig. 4A). Similarly, z-scored spike counts were also significantly greater in response to unexpected gains than to unexpected losses [$F_{1,14} = 18.2$, MSE = 24, $P < 0.001$] (Fig. 4B). This difference in spike activity was driven by a statistically significant response to unexpected gains greater than baseline activity (SOM text).

We confirmed that this difference in aggregated spike activity was consistently observed in individual cells by examining the relative differences in spike activity in response to unexpected gains and losses for each cell. During this interval, significantly more cells [14 out of 15 cells; $\chi^2(1) = 11.27$, $P < 0.001$] exhibited higher normalized spike rates in response to unexpected gains than to unexpected losses [mean difference of 0.67 ± 0.14 (mean \pm SEM) z-scored spikes per second].

To confirm that human SN activity is primarily responsible for differentiating only between unexpected gains and losses, we examined differences in spiking activity between expected

gains and losses. In the same 225-msec post-feedback interval, the difference in spike rates and normalized spike counts between expected gains and losses did not approach significance [$F_{1,14} < 1$, n.s.] (Fig. 4C). Similarly, the remaining intervals exhibited no significant differences in spike rate or normalized spike count in response to expected gains and losses.

The computation of how outcomes differ from expectation, often referred to as prediction error (2), is a central component of models of reinforcement learning and thought to be encoded by the activity of dopaminergic neurons (5–7, 15, 26). We examined the correlation between spike activity and changes in expected reward as determined by Eq. 1 under the assumption that this change can be used as a surrogate for prediction error. We defined prediction error here as the trial-to-trial adjustment each participant makes to the expected reward for each deck as determined by our model of expectation. Mean spike rates in the same post-feedback interval during trials associated with large positive prediction errors were larger than spike rates associated with small positive prediction errors, but this difference was only marginally significant [$F_{1,14} = 3.2$, MSE = 8, $P = 0.09$] (20). As trial-to-trial differences in expected reward in-

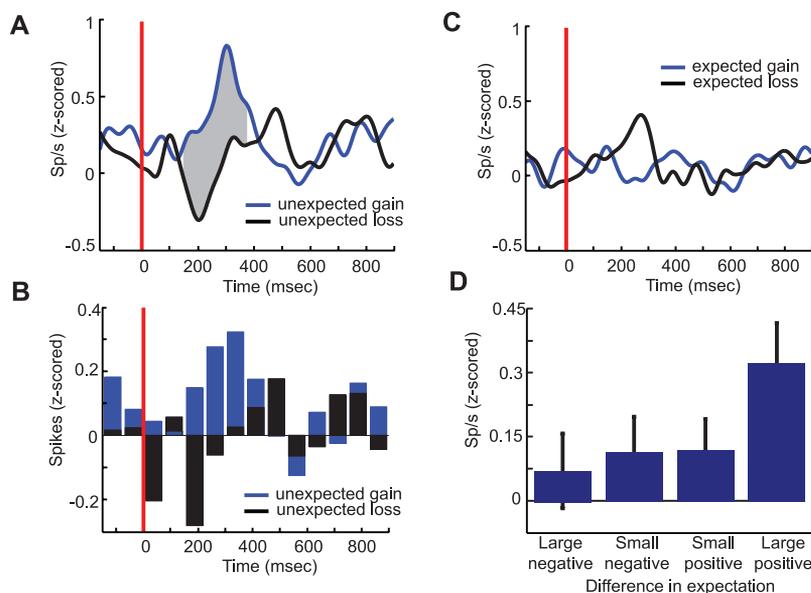


Fig. 4. (A) Average z-scored spike rate for unexpected gains (blue trace) compared with unexpected losses (black trace). The red line indicates feedback onset. The gray shaded region indicates the 225-msec interval between 150 and 375 msec after feedback onset. Traces represent average activity from 15 SN cells recorded from 10 participants. (B) Average z-scored spike histograms for unexpected gains (blue bars) compared to unexpected losses (black bars). The red vertical line indicates feedback onset. Histograms represent average z-scored spike counts from the same 15 SN cells. (C) Average z-scored spike rate for expected gains (blue trace) did not differ significantly from expected losses (black trace) for any interval. The red line indicates feedback onset. (D) For every participant, the median positive and negative trial-to-trial change in expected reward, as determined by Eq. 1, is used to classify prediction error into large and small positive and negative differences. Mean z-scored spike rate, captured between 150 and 375 msec after feedback onset for all cells, is shown for each level of prediction error. Error bars represent SEM.

creased, there was a general increase in spike activity, but this trend was also only marginally significant (Fig. 4D) (20).

Our results show that differences in human SN responses to positive and negative feedback are mainly driven by unexpected outcomes, with no significant differences in neural activity for outcomes that are anticipated according to our model. By responding to unexpected financial rewards, these putatively dopaminergic cells encode information that probably helps participants maximize reward in the probabilistic learning task.

Our results address the important question of whether extrapolating findings about the reward properties of dopaminergic SN neurons from nonhuman primates to humans is reasonable (27). Whereas the role of midbrain dopaminergic neurons in reward learning has been studied extensively in animals (4–8, 15, 26), the evidence presented here represents direct measurement of SN neurons in humans who were engaged in a probabilistic learning task. Our findings should serve as a point of validation for animal models of reward learning.

The reward for choosing the correct deck in our study was a perceptual stimulus designed to evoke a cognitive representation of financial reward. Primate studies, which often rely on highly salient first-order reward stimuli

such as food and water, have demonstrated that dopaminergic neurons are also capable of responding to second-order associations (28), which are items that can be used to directly satisfy first-order needs. Because no monetary compensation was directly provided, our abstract rewards (i.e., images of second-order rewards) may be considered third-order. That the modest third-order rewards used here elicited a significant dopaminergic response, when they were unexpected, suggests that SN activity may play a more widespread role in reinforcement learning than was previously thought.

Our findings suggest that neurons in the human SN play a central role in reward-based learning, modulating learning based on the discrepancy between the expected and the realized outcome (1, 2). These findings are consistent with the hypothesized role of the basal ganglia, including the SN, in addiction and other disorders involving reward-seeking behavior (29). More importantly, these findings are consistent with models of reinforcement learning involving the basal ganglia, and they suggest a neural mechanism underlying reward learning in humans.

References and Notes

1. R. A. Rescorla, A. R. Wagner, *Classical Conditioning II: Current Research and Theory* (Appleton Century Crofts, New York, 1972), pp. 64–99.

2. R. Sutton, A. Barto, *Learning and Computational Neuroscience: Foundations of Adaptive Networks* (MIT Press, Cambridge, MA, 1990), pp. 497–437.
 3. P. R. Montague, P. Dayan, T. J. Sejnowski, *J. Neurosci.* **16**, 1936 (1996).
 4. J. Mirenovic, W. Schultz, *Nature* **379**, 449 (1996).
 5. W. Schultz, P. Dayan, P. R. Montague, *Science* **275**, 1593 (1997).
 6. W. Schultz, *Neuron* **36**, 241 (2002).
 7. J. R. Hollerman, W. Schultz, *Nat. Neurosci.* **1**, 304 (1998).
 8. M. A. Ungless, P. J. Magill, J. P. Bolam, *Science* **303**, 2040 (2004).
 9. H. M. Bayer, P. W. Glimcher, *Neuron* **47**, 129 (2005).
 10. S. M. McClure, D. I. Laibson, G. Loewenstein, J. D. Cohen, *Science* **306**, 503 (2004).
 11. S. C. Tanaka *et al.*, *Nat. Neurosci.* **7**, 887 (2004).
 12. A. M. Graybiel, *Curr. Opin. Neurobiol.* **15**, 638 (2005).
 13. M. J. Frank, *J. Cogn. Neurosci.* **17**, 51 (2005).
 14. N. Daw, K. Doya, *Curr. Opin. Neurobiol.* **16**, 199 (2006).
 15. G. Morris, A. Nevet, D. Arkadir, E. Vaadia, H. Bergman, *Nat. Neurosci.* **9**, 1057 (2006).
 16. B. J. Knowlton, J. A. Mangels, L. R. Squire, *Science* **273**, 1399 (1996).
 17. R. Cools, R. Barker, B. J. Sahakian, T. W. Robbins, *Cereb. Cortex* **11**, 1136 (2001).
 18. M. J. Frank, L. Seeberger, R. C. O'Reilly, *Science* **306**, 1940 (2004); published online 4 November 2004 (10.1126/science.1102941).
 19. R. Cools, R. Barker, B. J. Sahakian, T. W. Robbins, *Neuropsychologia* **41**, 1431 (2003).
 20. Materials and methods are available as supporting material on Science Online.
 21. J. Wixted, E. Ebbesen, *Psychol. Sci.* **2**, 409 (1991).
 22. D. Rubin, A. Wenzel, *Psychol. Rev.* **103**, 734 (1996).
 23. For every trial, $E_t[n]$ represents the current expected reward for the left deck. If $E_t[n]$ exceeds 0.5 for a particular trial n , and if choosing the left deck on that trial yields positive feedback, we label this event an expected win. Conversely, if $E_t[n]$ is below 0.5 for a trial that yields positive feedback after choosing the left deck, we label this event an unexpected win. Similarly, we define expected and unexpected losses as those cases when negative feedback occurs when $E_t[n]$ is below or above 0.5, respectively. For trials when the right deck is chosen, we use $E_r[n]$ in identical fashion to classify responses into one of these four categories.
 24. W. Schultz, A. Ruffieux, P. Aebischer, *Exp. Brain Res.* **51**, 377 (1983).
 25. W. Schultz, *J. Neurophysiol.* **56**, 1439 (1986).
 26. H. Nakahara, H. Itoh, R. Kawagoe, Y. Takikawa, O. Hikosaka, *Neuron* **41**, 269 (2004).
 27. C. D. Hardman *et al.*, *J. Comp. Neurol.* **445**, 238 (2002).
 28. W. Schultz, P. Apicella, T. Ljunberg, *J. Neurosci.* **13**, 900 (1993).
 29. S. E. Hyman, R. C. Malenka, E. J. Nestler, *Annu. Rev. Neurosci.* **29**, 565 (2006).
 30. This work is partially supported by NIH grant MH61975, Conte Center grant MH062196, and the Dana Foundation. We thank the staff of the Pennsylvania Neurological Institute for their assistance. We also thank J. Wachter and A. Krieger for statistical advice, A. Geller for cognitive task programming, and M. Kilpatrick and P. Connolly for assistance with DBS surgery.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5920/1496/DC1
 Materials and Methods
 SOM Text
 References

17 October 2008; accepted 16 January 2009
 10.1126/science.1167342

ERRATUM

Post date 24 July 2009

Reports: "Human substantia nigra neurons encode unexpected financial rewards" by K. A. Zghloul *et al.* (13 March, p. 1496). Equation 1 in this paper was incorrect. The correct equation is as follows: $E_d[n] = 0.5 + 0.5 \sum_{i=1}^{n-1} R_d[n-i] \alpha i^{-\tau}$ $n = 2, \dots, N$.