Behavioral/Cognitive

Microstimulation of the Human Substantia Nigra Alters Reinforcement Learning

Ashwin G. Ramayya,[1] Amrit Misra,[4] Gordon H. Baltuch,[3]* and Michael J. Kahana[2]*

[1]Neuroscience Graduate Group, [2]Department of Psychology, and [3]Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19106, and [4]Drexel University School of Biomedical Engineering, Science and Health Systems, Philadelphia, Pennsylvania 19103

Animal studies have shown that substantia nigra (SN) dopaminergic (DA) neurons strengthen action–reward associations during reinforcement learning, but their role in human learning is not known. Here, we applied microstimulation in the SN of 11 patients undergoing deep brain stimulation surgery for the treatment of Parkinson's disease as they performed a two-alternative probability learning task in which rewards were contingent on stimuli, rather than actions. Subjects demonstrated decreased learning from reward trials that were accompanied by phasic SN microstimulation compared with reward trials without stimulation. Subjects who showed large decreases in learning also showed an increased bias toward repeating actions after stimulation trials; therefore, stimulation may have decreased learning by strengthening action–reward associations rather than stimulus–reward associations. Our findings build on previous studies implicating SN DA neurons in preferentially strengthening action–reward associations during reinforcement learning.

Key words: **dopamine; human; microstimulation; Parkinson's disease; reinforcement learning; substantia nigra**

Introduction

Contemporary theories of reinforcement learning posit that decisions are modified based on a reward prediction error (RPE), the difference between the experienced and predicted reward (Sutton and Barto, 1990). A positive RPE (outcome better than expected) strengthens associations between the reward and preceding events (e.g., stimuli, actions) such that a rewarded decision is more likely to be repeated. Animal electrophysiology studies have shown that dopaminergic (DA) neurons in the ventral tegmental area (VTA) and substantia nigra (SN) display phasic bursts of activity following unexpected rewards (Schultz et al., 1997; Bayer and Glimcher, 2005), leading to the hypothesis that they encode positive RPEs (Glimcher, 2011). Because SN DA neurons predominantly send projections to dorsal striatal regions that mediate action selection (Haber et al., 2000; Lau and Glimcher, 2008), they have been hypothesized to preferentially strengthen action–reward associations during reinforcement learning (Montague et al., 1996). Supporting this hypothesis, a previous rodent study has shown that SN microstimulation reinforces actions and strengthens corticostriatal synapses in a dopamine-dependent manner (Reynolds et al., 2001).

In humans, much of the evidence linking DA activity to reinforcement learning has come from studies in patients with Parkinson's disease (PD), who have significant degeneration of SN DA neurons (Ma et al., 1996) and show specific deficits on reward-based learning tasks compared with age-matched controls (Knowlton et al., 1996). Administration of DA agonists in these patients improves reinforcement learning performance (Frank et al., 2004; Rutledge et al., 2009), suggesting that DA plays an important role in human reinforcement learning. However, both PD and DA agonists manipulate tonic DA levels throughout the brain in addition to phasic DA responses. Because altered tonic DA levels may influence performance on learning tasks through nonspecific changes in motivation (Niv et al., 2007), these studies do not specifically implicate the phasic activity of DA neurons in human reinforcement learning (Shiner et al., 2012).

To study the role of phasic DA activity during human reinforcement learning, we applied microstimulation in the SN of patients undergoing deep brain stimulation (DBS) surgery for the treatment of PD. Microstimulation has been shown to enhance neural responses near the electrode tip (Histed et al., 2009) and is widely used in animal electrophysiology studies to map causal relations between particular neural populations and behavior (Clark et al., 2011). Although microstimulation is often applied during DBS to aid in clinical targeting (Lafreniere-Roula et al., 2009), it has not been applied in association with a cognitive task. Here, we applied microstimulation during a subset of feedback trials as subjects performed a reinforcement learning task in which rewards were contingent on stimuli rather than on actions (putative DA neurons in the human SN have been shown to

Received Dec. 29, 2013; revised March 25, 2014; accepted March 28, 2014.

Author contributions: A.G.R., G.H.B., and M.J.K. designed research; A.G.R. and G.H.B. performed research; A.M. contributed unpublished reagents/analytic tools; A.G.R. analyzed data; A.G.R. and M.J.K. wrote the paper.

This work was supported by the National Institutes of Health (Grant MH55687). We thank the subjects who volunteered in this study; M. Kerr, H. Chaibainou, and the staff of the Pennsylvania Neurological Institute for their help in coordinating patient recruitment for this study; M. Donley-Fletcher and M. Kinslow for assistance with intraoperative data collection; and K.A. Zaghloul, J.F. Burke, M.J.K. Healey, N.M. Long, and M.B. Merkow for insightful comments on this manuscript and general discussion.

The authors declare no competing financial interests.

*G.H.B. and M.J.K. contributed equally to this work.

Correspondence should be addressed to either of the following: Gordon H. Baltuch, Department of Neurosurgery, Perelman School of Medicine 235 South 8th Street, Philadelphia, PA 19104, E-mail: baltuchg@mail.med.upenn.edu; or Michael J. Kahana, Department of Psychology, University of Pennsylvania, 3401 Walnut Street, Room 303C, Philadelphia, PA 19106, E-mail: kahana@psych.upenn.edu.

DOI:10.1523/JNEUROSCI.5445-13.2014

 0270-6474/14/346887-09$15.00/0

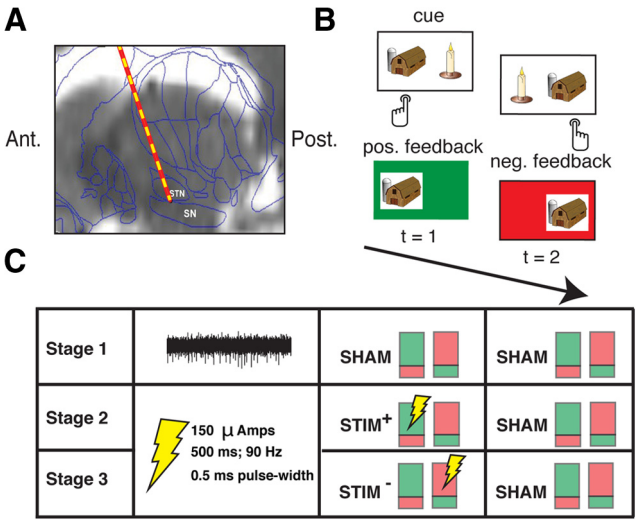

Figure 1. ***A***, Intraoperative targeting of SN. During DBS surgery, a microelectrode is advanced into the SN to map the ventral border of the STN. An example preoperative MRI scan (sagittal view) overlaid with a standard brain atlas and estimated microelectrode position is shown (Jaggi et al., 2004; Zaghloul et al., 2009). ***B***, Reinforcement learning task. During surgery, 11 subjects performed a two-alternative probability learning task with inconsistent stimulus-response mapping. ***C***, Experimental design. During each stage of the session (50 trials each), subjects sampled reward probabilities of two item pairs that were matched in relative reward rate. Each pair of colored rectangles depicts an item pair (the green and red shading within each rectangle indicates the probability of positive and negative feedback associated a particular item in the pair, respectively). During Stage 1, we obtained microelectrode recordings from the SN. An example 500 ms high-pass-filtered (>300 Hz) voltage trace is shown. During Stages 2 and 3, we applied electrical microstimulation through the recording microelectrode as depicted, but no longer obtained recordings (see Materials and Methods).

display RPE-like responses during this postfeedback time interval; Zaghloul et al., 2009). If phasic SN responses preferentially strengthen action–reward associations during reinforcement learning, then stimulation during reward trials should induce a bias to repeating actions, rather than stimuli, and disrupt learning during the task.

Materials and Methods

Subjects. Eleven patients undergoing DBS surgery for the treatment of PD volunteered to take part in this study (8 male, 3 female, age = 63 ± 7 years, mean ± SD). Subjects provided their informed consent during preoperative consultation and received no financial compensation for their participation. Per routine clinical protocol, Parkinson's medications were stopped on the night before surgery (12 h preoperatively); therefore, subjects engaged in the study while in an OFF state. The study was conducted in accordance with a University of Pennsylvania Institutional Review Board-approved protocol.

Intraoperative methods. During surgery, intraoperative microelectrode recordings (obtained from a 1-μm-diameter tungsten tip electrode advanced with a power-assisted microdrive) were used to identify the SN and the subthalamic nucleus (STN) as per routine clinical protocol (Jaggi et al., 2004; Fig. 1*a*). Electrical microstimulation is routinely applied through the microelectrode to aid in clinical mapping of SN and STN neurons and was approved for use in this study by the University of Pennsylvania Institutional Review Board. Once the microelectrode was positioned in the SN, we administered a two-alternative probability learning task through a laptop computer placed in front of the subject. Subjects viewed the computer screen through prism glasses placed over the stereotactic frame and expressed choices by pressing buttons on handheld controllers placed in each hand.

Reinforcement learning task. Subjects performed a two-alternative probability learning task with feedback, which has been widely applied to the study of reinforcement learning (Sugrue et al., 2005; Fig. 1*b*). Subjects chose between pairs of items and probabilistically received positive or negative feedback after each choice. One item in each pair was associated with a high probability of reward (e.g., 0.8), whereas the other item was associated with a low probability of reward (e.g., 0.2). Subjects were informed that each stimulus in a presented pair was associated with a distinct reward rate and that their goal was to maximize rewards over the entire session. To achieve this goal, subjects needed to learn the underlying reward probabilities associated with each stimulus by trial and error and adjust their choices accordingly. Each trial consisted of the presentation of stimuli, subject choice, and feedback presentation. In the event of positive feedback ("wins"), the screen turned green and the sound of a cash register was presented. In the event of negative feedback ("losses"), the screen turned red and an error tone was presented. The item pairs consisted of colored images of simple objects that were matched based on normative data (e.g., semantic similarity, naming agreement, familiarity, and complexity; Rossion and Pourtois, 2004). The same pairs of stimuli were used across subjects; however, the assignment of reward probabilities to each stimulus in the pair was randomly assigned for each subject. The arrangement of the items on the screen, and thus the button associated with each item (left and right), was randomized from trial to trial.

Each session consisted of 150 trials (15 min of testing time) and was subdivided into three stages (50 trials each; Fig. 1*c*). Each stage consisted of two novel pairs of stimuli (two sets of stimuli) that resulted in two independent learning conditions per stage. Such a design was used so that we could study the effects of stimulation on learning while controlling for various extraneous factors that might influence performance. To ensure a fair comparison between the two item pairs within each stage, the relative reward rates for each pair were set to 0.8 versus 0.2. If the subject selected the high-probability item on at least 80% of trials on Stage 1, then the relative reward rates for both pairs in subsequent stages were set to 0.7 versus 0.3 to encourage learning during the remainder of the session; otherwise, they remained the same. Furthermore, the item pairs were presented in alternating trains of three to six trials. This method of item presentation allowed subjects to learn reward probabilities associated with a single item pair for multiple sequential trials while ensuring that the two pairs within a stage were associated with similar levels of motivation, or arousal, which likely vary slowly throughout the session.

During Stage 1, we did not provide stimulation in association with either pair, but during the subsequent stages, we applied microstimulation during a subset of feedback trials (see "Stimulation Parameters" section). During Stage 2, one of the pairs was associated with SN microstimulation during positive feedback that followed a high-reward-probability choice ($STIM^+$), whereas the other pair did not receive stimulation ($SHAM^+$). During Stage 3, one pair received SN microstimulation during negative feedback that followed a low-reward-probability choice ($STIM^-$), whereas the other pair did not receive stimulation ($SHAM^-$). During Stage 2, we sought to assess the effect of stimulation on learning from wins by comparing performance on the $STIM^+$ and $SHAM^+$ pairs, whereas during Stage 3, we sought to assess the effect of stimulation on learning from losses by comparing performance on the $STIM^-$ and $SHAM^-$ pairs.

Because the goal of the study was to assess whether there were stimulation-related changes in learning across the various item pairs, it was crucial to minimize within-subject, across-pair variability in choice behavior. To reduce such variability, we ensured that reward probabilities of the items did not drastically fluctuate over the course of each stage by using deterministic reward schedules (e.g., for a reward probability of 0.8, we ensured that 4 of every 5 selections of that stimulus result in positive feedback). These deterministic reward schedules were not true binomial processes and may have allowed for more distinct learning strategies than reward schedules typically used in probability learning tasks. However, by reducing within-subject variability in choice behavior, these schedules allowed us to detect more effectively stimulation-related changes in learning and to take full advantage of the rare clinical opportunity offered by this patient population. When possible, subjects first performed the task during preoperative consultation, but in all cases, the task was reviewed with subjects on the morning of surgery. Further instructions were provided before beginning the task intraoperatively. Subject #3 did not perform Stage 1 due to a technical difficulty during the experiment, but completed Stages 2 and 3 of the task (see Table 2). The design also included a fourth stage consisting of a $STIM^+$ and a $STIM^-$

pair to allow for a direct comparison between the two conditions; however, because only a subset of subjects ($n = 6$) completed this stage due to fatigue, these data were not analyzed for this study.

Stimulation parameters. Stimulation was provided through the microelectrode immediately after feedback presentation during the learning task using an FHC Pulsar 6b microstimulator using the following parameters: biphasic, cathode phase-lead pulses at 90 Hz, lasting 500 ms at an amplitude of 150 μAmps and a pulse width of 500 μs. Similar stimulation parameters have induced learning in the rodent SN (Reynolds et al., 2001) and the nonhuman primate VTA (Grattan et al., 2011). An LED on the front chasse of the stimulator indicated the onset of stimulation; however, this was not visible to the patient as they performed the task. There was no sound associated with stimulation. Therefore, stimulation trials were not signaled to subjects in any manner. None of the subjects reported a perceptual change associated with the application of microstimulation.

Reinforcement learning model simulations. To better understand subjects' behavior during the task, we simulated the performance of various reinforcement learning models (see Q learning model, below) on a two-alternative probability learning task with inconsistent stimulus–response mapping. Each simulated session consisted of 25 trials (similar to one item pair in our task) and consisted of a single item pair with reward probabilities of 0.8 and 0.2. Each item was randomly assigned to an action from trial to trial.

Q learning model. This standard reinforcement learning model maintains independent estimates of reward expectation (Q) values for each option i at each time t (Sutton and Barto, 1990). A choice is probabilistically generated on each trial by comparing the Q values of available options on that trial using the following logistic function: $P_i(t) = \frac{\exp(Q_i(t)/\beta)}{\sum_j \exp(Q_j(t)/\beta)}$. Where β is a free parameter for inverse gain in the softmax logistic function (which accommodates noise in the choice process or different relative tendencies for exploration vs exploitation; Daw et al., 2006). Once an item is selected by the model, feedback is received, and Q values are updated using the following learning rule:

$$Q_i(t+1) = Q_i(t) + \alpha[R(t) - Q_i(t)],$$

where $R(t) = 1$ for correct feedback, $R(t) = 0$ for incorrect feedback, and α is the learning rate parameter that adjusts the manner in which previous reinforcements influence current Q values. Large α values (upper bound = 1) heavily weight recent outcomes when estimating Q, whereas small α values (lower bound = 0) more evenly weight reinforcements from previous trials. To simulate the behavioral changes associated with decreasing learning rates, we studied the performance of 34 Q model agents that varied in their α values (0.01–1, with a step size of 0.03; Frank et al., 2007) while fixing the β parameter at 0.2. Similarly, to simulate behavioral changes associated with increasing noise in the choice policy, we studied the performance of 34 agents that varied in their β values (0.01–1, with a step size of 0.03) while fixing the α parameter at 0.2. Q values associated with each item were initialized to 0.5. We simulated the performance of these agents on 1000 randomly generated sessions.

Hybrid action-stimulus learning model. To extend the Q learning model to a task with inconsistent stimulus-response mapping, we developed a hybrid action-stimulus (AQ) learning model. Similar to the standard Q model, the hybrid AQ model tracks reward expectations associated with each stimulus using a recency-weighted exponential decay function that is controlled by the learning rate α (ranging from 0–1). However, in addition, the hybrid AQ model also tracks the reward expectations associated with each available action (A). To limit the addition of free parameters, the α associated with the action values is assumed to be the same for tracking stimulus and action values. A weighting parameter (W_A, ranging from 0 to 1) determines the aggregate reward expectation associated with a particular action/stimulus combination (AQ) in the following manner:

$$AQ_{i,j}(t) = W_A(A_i(t)) + (1 - W_A)(Q_j(t)),$$

where i indexes a particular stimulus, j indexes a particular action, and t represents a particular trial. Similar to the Q model, the hybrid AQ model computes the probability of selecting from each action/stimulus combination using the following softmax-logistic function: $P_{i,j}(t) = \frac{\exp(AQ_{i,j}(t)/\beta)}{\sum_j \exp(AQ_{i^*,j^*}(t)/\beta)}$, where AQ_{i^*,j^*} represents all other available action-stimulus combinations and β is a free parameter for inverse gain in the softmax logistic function. In summary, the hybrid AQ model has three free parameters—the learning rate (α), noise in the choice policy (β), and an action-value weighting parameter (W_A). To simulate the behavioral changes that would be observed after strengthened action–reward associations, we simulated the behavior of 34 hybrid AQ models at various levels of the W_A parameter (0.01–1, with a step size of 0.03) while fixing α and β at 0.2.

Fitting reinforcement learning models to subjects' behavioral data. To study the relation between stimulation-related behavioral changes and the parameters of the reinforcement learning models directly, we fit the two-parameter Q learning model and the three-parameter hybrid AQ model to each subject's behavioral data. We fit each model separately to subjects' choices on each item pair so as to compare changes in the parameter values across stimulation conditions. To identify the set of best-fitting parameters for a given pair, we performed a grid search through each model's parameter space (0.01–1, with a step size of 0.03) and selected the set of parameters that resulted in the most positive log-likelihood estimate (LLE) of the model's predictions of the subject's choices (i^*) as follows: $LLE = log\left(\prod_t P_{i^*,t}\right)$. To assess the goodness-of-fit of each model fit across the dataset, we computed a LLE of each model's predictions of all subject choices during each item pair. To assess whether model predictions were better than chance, we computed a pseudo-R^2 statistic (r-LLE)/r, where r represents the LLE of purely random choices ($P = 0.5$ for all choices; Daw et al., 2006). To allow for a fair comparison between the two- and three-parameter model fits, we penalized each model for complexity by using the Akaike Information Criterion (AIC; Akaike, 1974). Because we were computing goodness-of-fit on the group level, we considered the Q model to have 22 parameters (two parameters for each subject), and the hybrid AQ model to have 33 parameters (three parameters for each subject).

Extracting spiking activity from microelectrode recordings. We obtained microelectrode recordings as subjects performed Stage 1 before applying microstimulation during the experiment. Because these recordings were of a relatively short duration ($\approx$ 5 min.) and were only associated with 50 trials, their main purpose was to aid in interpretation of the stimulation results, rather than to characterize the functional properties of human SN neuronal activity (Zaghloul et al., 2009). To assess whether stimulation-related behavioral changes were related to the properties of the neuronal population near the electrode tip, we extracted multiunit activity from each microelectrode recording using the WaveClus software package (Quiroga et al., 2005). We band-pass filtered each voltage recording from 400 to 5000 Hz and manually removed periods of motion artifact. We identified spike events as negative deflections in the voltage trace that crossed a threshold that was manually defined for each recording ($\approx$3.5 SD about the mean amplitude of the filtered signal). The minimum duration between consecutive spike events (censor period) was set to be 1.5 ms. Spike events were subsequently clustered into units based on the first three principal components of the waveform. Noise clusters from motion artifact or power line contamination were manually invalidated. We considered spikes from all remaining clusters together as a multiunit. From each multiunit, we extracted two features that are characteristic of DA activity—the mean waveform duration and the phasic postreward response (Zaghloul et al., 2009; Ungless and Grace, 2012). We quantified the waveform duration as the mean peak-to-trough duration for all spikes and the phasic postreward response as the difference between the average spike rate during the 0–500 ms postreward interval and that during the −250–0 and 500–750 ms intervals. We did not consider responses after negative outcomes because DA neurons are not homogenous in their responses after negative outcomes (Matsumoto and Hikosaka, 2009). We obtained multiunit activity from nine of the 11

subjects. We were unable to obtain recordings from one subject (Subject #3) and could not distinguish spiking activity from noise contamination in another subject (Subject #11).

Results

We applied microstimulation in the SN of 11 patients undergoing DBS surgery for the treatment of PD (Fig. 1*a*). Subjects performed a two-alternative probability learning task in which they selected between pairs of items (images of common objects) and probabilistically received abstract rewards ("wins") or punishments ("losses") after each choice (Fig. 1*b*). Subjects were instructed that one item in each pair carried a higher reward probability than the other item in the pair and that their goal was to maximize the number of rewards that they obtained during the session. We indexed learning on a given item pair by calculating the probability that subjects selected the high-probability item on trials associated with that pair. Because items were randomly assigned to an action (left or right button) on each trial, subjects were required to encode stimulus–reward associations rather than action–reward associations to perform well during the task. The task was divided into multiple stages (50 trials each) with each stage consisting of two item pairs matched in their relative reward rates (see Materials and Methods; Fig. 1*c*). During Stage 1, we did not provide stimulation in association with either item pair (SHAM) so that subjects could become acclimated to the learning task. Across the 50 trials of Stage 1, subjects selected the high-probability item on 63% of trials, which trended toward being greater than chance (50%, $t(9) = 2.07, p = 0.068$). In each of the next two stages, one item pair was associated with microstimulation (STIM), whereas the other was not (SHAM). By comparing learning on the STIM and SHAM pair within each stage, we sought to assess the effects of SN microstimulation on learning.

During Stage 2, we assessed the effect of stimulation on reward learning by applying stimulation during positive outcomes associated with the high-reward-probability item on one of the pairs ($STIM^+$). We found that subjects were less likely to select the high-probability item on the $STIM^+$ pair compared with the SHAM pair during this stage ($t(10) = 2.56$, $p = 0.029$; Fig. 2, Table 1). This difference in performance could be attributed to a stimulation-related decrease in learning; subjects demonstrated learning on the SHAM pair (accuracy = 67%, $t(10) = 3.05$, $p = 0.012$), but did not perform better than chance on the $STIM^+$ pair (accuracy = 48%, $p > 0.5$). To directly study the behavioral changes that occured following stimulation, we compared subjects' tendencies to repeat their selection of the high-reward-probability item after rewards ("win-stay") on the $STIM^+$ and the SHAM pair. We found that subjects reliably demonstrated decreased win-stay after reward trials accompanied by stimulation compared with reward trials without stimulation ($t(10) = 2.71$, $p = 0.022$). Therefore, subjects demonstrated decreased learning from reward trials that were accompanied by phasic SN microstimulation compared with reward trials without stimulation. During Stage 3, we applied stimulation during negative feedback associated with the low-reward probability item on one item pair ($STIM^-$) to study the effect of SN stimulation on learning from negative outcomes. We did not observe differences in learning between the $STIM^-$ pair and the SHAM pair within the same stage, either in terms of overall accuracy (Fig. 2) or their probability repeating an item choice after stimulation trials ($p's > 0.3$).

Our main finding was that SN microstimulation after rewards during Stage 2 disrupted learning of stimulus–reward associations. Because SN DA neurons have been hypothesized to preferentially strengthen action–reward associations (Montague et al., 1996; Haber et al., 2000; Frank and Surmeier, 2009), the observed decrease in learning might have occurred because stimulation induced a bias toward repeating actions rather than stimuli after high-probability reward trials. Such a bias would result in decreased performance because the mapping between stimuli and actions (left vs right button) was randomized from trial to trial during the task; repeating the same action after the selection of a high reward-probability item would result in the selection of the low-reward-probability item on approximately half the trials. If this is the case, then subjects should show an increased bias toward repeating the same button after high-probability-reward trials ("win-same button") on the $STIM^+$ pair compared with the SHAM pair. We did not observe a reliable stimulation-related increase in win-same button across subjects ($p > 0.4$); however, we observed a positive correlation between stimulation-related decreases in accuracy and increases in win-same button ($r = 0.77$, $p = 0.006$; Fig. 3*a*). Therefore, subjects who showed the greatest stimulation-related decreases in learning also showed an increased bias toward repeating actions after stimulation trials.

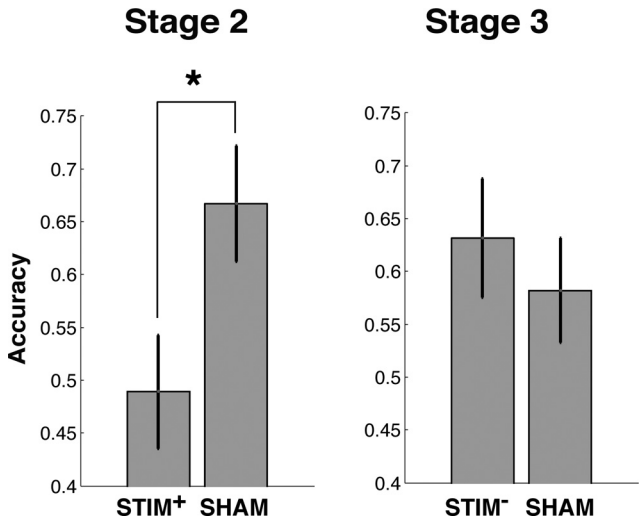

Figure 2. Effects of stimulation on learning. To index learning performance on a particular item pair, we computed the probability that subjects chose the item that was associated with a high reward probability ("accuracy"). During Stage 2, subjects demonstrated lower accuracy on the $STIM^+$ pair compared with the SHAM pair. During Stage 3, we did not identify changes in accuracy between the $STIM^-$ and SHAM pairs. $*p < 0.05$. Error bars indicate SEM across subjects ($n = 11$).

The positive correlation between stimulation-related decreases in accuracy and increases in win-same button suggests that stimulation may have disrupted learning by strengthening action–reward associations during the task. However, one might wonder whether this positive correlation might simply occur in association with decreased learning during our task. To assess whether this was the case, we simulated the performance of a standard two-parameter reinforcement learning model (*Q*-model; Sutton and Barto, 1990) performing a two-alternative probability learning task with inconsistent stimulus-response mapping (Materials and Methods; Fig. 3*b*,*c*). Briefly, the model estimates the expected reward associated with each stimulus based on a recency-weighted average of recent outcomes (forgetting function) and probabilistically makes a selection by comparing the expected reward associated with the available options. The model has two free parameters: a learning rate (α) that controls the rate

Table 1. Summary of subject data

Subject	Age	Sex	Δ Accuracy	Δ Win-stay	Δ Win-same button	Waveform duration (ms)	Phasic spike response (sp/s)
1	67	M	+0.12	−0.50	−0.17	0.77	−1.13
2	66	M	−0.36	−0.17	+0.21	0.78	0.34
3	66	M	−0.16	+0.025	−0.17	—	—
4	53	F	+0.08	+0.028	0	0.75	1.36
5	74	M	−0.32	−0.52	+0.20	0.84	−0.86
6	54	M	−0.68	−1.00	+0.53	0.85	2.07
7	56	M	−0.28	−0.67	+0.17	0.85	1.07
8	68	M	+0.04	−0.13	−0.29	0.73	−0.73
9	53	M	−0.08	0	+0.33	0.92	1.43
10	61	F	−0.20	−0.03	−0.03	0.87	0.57
11	66	F	−0.12	−0.13	−0.13	—	—

Columns 4–6 describe behavioral changes during Stage 2; columns 7–8 describe properties of multiunit activity recorded during Stage 1.

—, Missing data (we were unable to obtain recordings from Subject #3 and did not identify spiking activity from Subject #11).

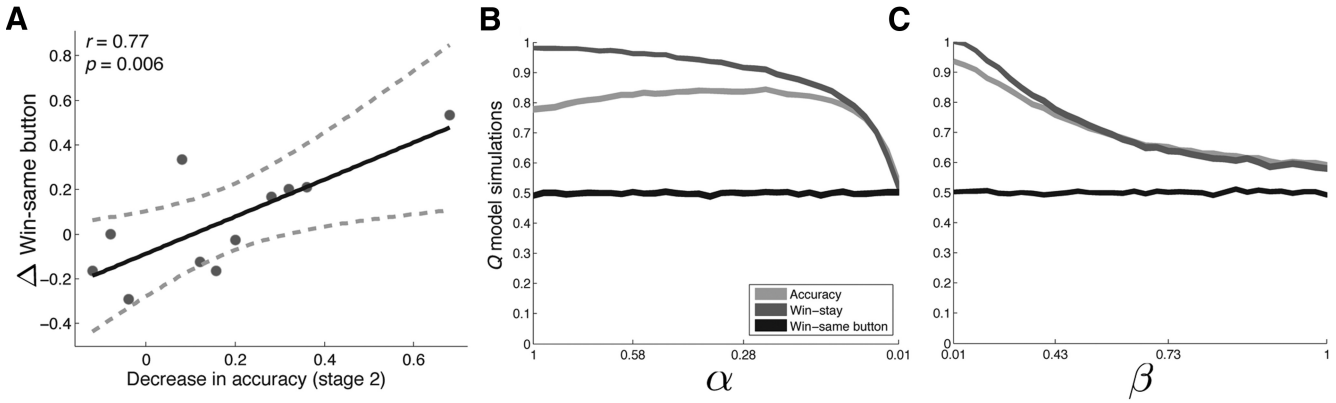

Figure 3. ***A***, Relation between decreases in learning and action bias. Stimulation-related decreases in accuracy were positively correlated with an increased bias toward repeating a button press after reward trials (win-same button; Pearson's $r = 0.77$, $p = 0.006$). Each dot represents a subject, the solid black line is the regression slope, and the dashed lines represent 95% confidence intervals. ***B***, ***C***, The Q learning model is insufficient to explain stimulation-related behavioral changes. Simulated behavior of a standard two-parameter reinforcement learning algorithm (Q model) on a two-alternative probability learning task with inconsistent stimulus-response mapping. Accuracy (light gray line), probability of repeating rewarded items (win-stay, dark gray line), and probability of repeating rewarded actions (win-same button, black line) are shown for decreasing learning rates (α; ***B***) and increasing noise in the choice policy (β; ***C***). Decreases in learning rate and increases in decision noise were accompanied by a decrease in accuracy and a decrease in win-stay, but no change in win-same button.

of decay of the forgetting function and noise in the choice policy (β). We found that both decreases in α and increases in β were associated with decreases in accuracy and win-stay, but no accompanying change in win-same button. Therefore, the positive correlation between decreased accuracy and increased win-same button cannot be explained by parametric changes in the standard two-parameter Q-model and is not a necessary result of the task design.

To assess whether the observed stimulation-related behavioral changes could be explained by strengthened action–reward associations, we developed a hybrid AQ learning algorithm that independently tracks reward expectations associated with each available action in addition to those associated with each available stimulus (see Materials and Methods). The model selects between available options by comparing the aggregate reward expectancies associated with the available action/stimulus combinations (e.g., house and left button press vs candle and right button press). A weighting parameter (W_A) controls the strength of action value representations relative to stimulus value representations (higher W_A values result in strengthened action–reward associations). In total, the model has three free parameters—α (the learning rate), β (noise in the choice policy), and W_A (strength of action–reward associations). We studied the behavior of the hybrid AQ model at various levels of W_A to simulate the behavioral changes that would be observed after strengthened action–reward associations (see Materials and Methods; Fig. 4*a*). We found that increasing levels of W_A were associated with decreased accuracy, decreased win-stay, and an increased win-same button. Therefore, increasing the strength of action–reward associations in the hybrid AQ model is able to explain the major stimulation-related behavioral changes, including the positive correlation between decreases in accuracy and increases in win-same button. Consistent with the behavior predicted by these model simulations, the five subjects who showed stimulation-related increases in win-same button showed a mean (±SEM) win-same button of 0.77 (±0.11) during the STIM$^+$ condition and 0.48 (±0.11) during the SHAM condition.

To investigate directly whether stimulation-related behavioral changes were related to strengthened action–reward associations, we fit the two-parameter Q model and the three-parameter hybrid AQ model to each subjects' choice behavior during the STIM$^+$ and SHAM conditions (see Materials and Methods). For each subject, we identified the parameter sets that provided the best fit to subjects' choices during each pair using a grid search across each model's parameter space. We assessed whether the three-parameter hybrid AQ model provided a better explanation of subjects' choice behavior than the two-parameter Q learning model using the AIC, a goodness-of-fit measure that applies a penalty for model complexity (Akaike, 1974). We found that the hybrid AQ model provided a better fit to subjects' choice behavior during the STIM$^+$ condition, whereas the $Q-$model provided a better fit to subjects' choice behavior during the

SHAM condition (Table 2). Then, using the parameter estimates obtained from the hybrid AQ model, we assessed whether stimulation-related decreases in accuracy during Stage 2 were best explained by changes in α, β, or W_A by applying the following linear regression model: $R = \beta_0 + \beta_A A + \beta_B B + \beta_W W$, where R was a vector containing the decrease in accuracy for each subject. A, B, and W were vectors containing changes in α, β, and W_A, respectively. We found that simulation-related decreases in accuracy demonstrated a significant, positive relation with increases in W_A ($\beta_W = 0.22, t(10) = 2.48, p = 0.017$), but not with changes in α or β (p's > 0.3). These results provide further support for the hypothesis that stimulation-related decreases in accuracy were related to strengthened action–reward associations.

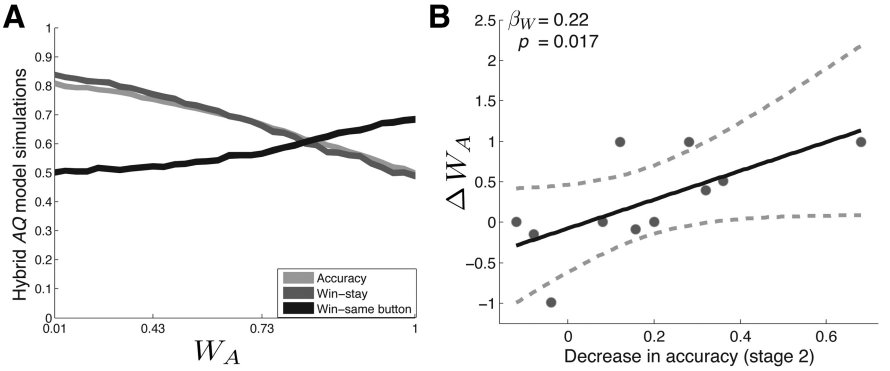

Figure 4. ***A***, Hybrid AQ learning model. Shown is the simulated behavior of the three-parameter reinforcement learning algorithm (hybrid AQ model) on a two-alternative probability learning task with inconsistent stimulus-response mapping. Accuracy (light gray line), probability of repeating rewarded items (win-stay, dark gray line), and probability of repeating rewarded actions (win-same button, black line) are shown for varying values of the action value weighting parameter (W_A). Strengthened action–reward associations were associated with decreases in accuracy and win-stay and increases in win-same button. ***B***, Stimulation-related behavioral changes can be explained by strengthened action–reward associations. We quantitatively fit the hybrid AQ model to subjects' behavior on the STIM$^+$ and SHAM pair during Stage 2. We found that stimulation-related decreases in accuracy showed a significant positive relation with increases in W_A, but not α or β. See main text for statistics.

Strengthened action–reward associations during feedback trials should result in improved accuracy during congruent trials (where the rewarded item is associated with the same action as the previous trial), but decreased accuracy during incongruent trials (where the rewarded item is no longer associated with the same action as the previous trial). Our finding that increases in win-same button were correlated with decreases in accuracy suggests that strengthened action–reward associations may have preferentially occurred during incongruent trials. To assess whether this was the case, we studied raw probabilities of win-same button during the SHAM and STIM$^+$ pairs in subjects who showed a stimulation-related increase in win-same button, but separately for congruent and incongruent trials ($n = 5$; Fig. 5*a*). During incongruent trials, these subjects showed a mean win-same button of 0.75 ($\pm$0.19) during the STIM$^+$ condition, but a win-same button of 0.24 ($\pm$0.15) during the SHAM condition. However, during congruent trials, these subjects showed a mean win-same button of 0.67 ($\pm$0.21) and 0.87 ($\pm$0.08) during the STIM$^+$ and SHAM conditions, respectively. To relate these behavioral patterns to the earlier model-based analyses, we examined the predicted win-same button probabilities of the various model simulations during congruent and incongruent trials. We found that the predictions of the Q learning model were inconsistent with the observed behavior because both decreases in α and increases in β were associated with symmetric changes in win-same button (decreases during congruent trials and increases during incongruent trials to chance level; Fig. 5*b*,*c*). In contrast, increases in W_A of the hybrid AQ model were associated with asymmetric changes in win-same button (increases in win-same button during incongruent trials to above chance levels and modest decreases in win same-button during congruent trials; Fig. 5*d*), which is consistent with the observed stimulation-related behavioral changes. One might have predicted that strengthened action–reward associations should result in increased win-same button after congruent trials; however, because each action is associated with a reward probability of 0.5, this would only occur in the setting of very high α values.

These results suggest that stimulation may have strengthened action–reward associations during the task, possibly by enhancing phasic DA activity in the SN (Montague et al., 1996; Reynolds et al., 2001). Because DA neurons are anatomically clustered in the SN (Henny et al., 2012) and because microstimulation has been shown to enhance the activity of neurons that surround the electrode tip (Histed et al., 2009), one might expect to observe the greatest changes in win-same button when the microelectrode tip was positioned near DA neurons. Therefore, we studied the relation between stimulation-related changes in win-same button and the properties of the neural activity recorded from the microelectrode during Stage 1. We extracted multiunit spiking activity from each recording and extracted two features that are characteristic of DA activity—average waveform duration and the phasic postreward response (Zaghloul et al., 2009; Ungless and Grace, 2012; see Materials and Methods). We found positive correlations between stimulation-related increases in win-same button and both the phasic postreward response (Pearson's $r = 0.69$, $p = 0.040$; Fig. 6*a*) and the mean waveform duration of recorded multiunit activity (Pearson's $r = 0.66$, $p = 0.053$; Fig. 6*b*). Multiunits recorded from the two subjects that showed the greatest increases in win-same button showed broad waveforms (0.85 and 0.92 ms) and phasic postreward bursts that were visible in the spike raster (+2.07 spikes/s, and +1.43 spikes/s; Fig. 6*c*). These results suggest that stimulation-related increases in win-same button were greatest when the microelectrode was positioned near neural populations that displayed properties characteristic of DA neurons.

Discussion

We applied electrical microstimulation in SN of 11 patients undergoing DBS surgery for the treatment of PD as they performed a two-alternative probability learning task in which rewards were contingent on stimuli rather than actions. Subjects were required to learn stimulus–reward associations rather than action–reward associations to perform well on the task. We found that SN microstimulation applied during reward trials disrupted learning compared with a control learning condition.

Phasic SN activity is functionally important for human reinforcement learning

By showing that SN microstimulation during the phasic postreward interval alters performance during the task, our findings provide an important bridge between animal and human studies of reinforcement learning. Animal studies have shown that the phasic activity of DA neurons signal positive RPEs that

Table 2. Summary of reinforcement learning model fits

Condition	α	β	W_A	Hybrid AQ pseudo-R^2 (AIC)	Q pseudo-R^2 (AIC)
SHAM	0.30 (±0.12)	0.31 (±0.11)	0.47 (±0.14)	0.23 (369.7)	0.20 (361.3)
STIM$^+$	0.38 (±0.11)	0.44 (±0.11)	0.71 (±0.12)	0.14 (404.7)	0.07 (412.8)

Means (±SEM) are shown for best-fitting AQ model parameter values (columns 2–4) associated with the STIM$^+$ and SHAM pairs during Stage 2. We report pseudo-R^2 and AIC goodness-of-fit measures for the three-parameter hybrid AQ model (column 5) and the two-parameter Q model (column 6) for each condition (see Materials and Methods).

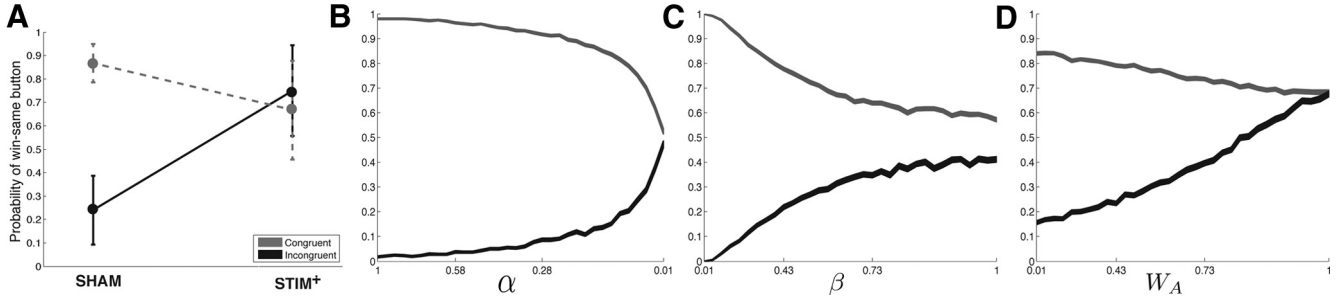

Figure 5. Win-same button during congruent and incongruent trials. ***A***, Subjects who showed stimulation-related increases in win-same button ($n = 5$) showed asymmetric changes during congruent (gray) and incongruent (black) trials when comparing STIM$^+$ and SHAM trials. ***B***, ***C***, Simulated behavior of a Q learning model shows symmetric changes in win-same button during congruent and incongruent trials. ***D***, Strengthened action–reward associations in the hybrid AQ learning model results in asymmetric changes in win-same button.

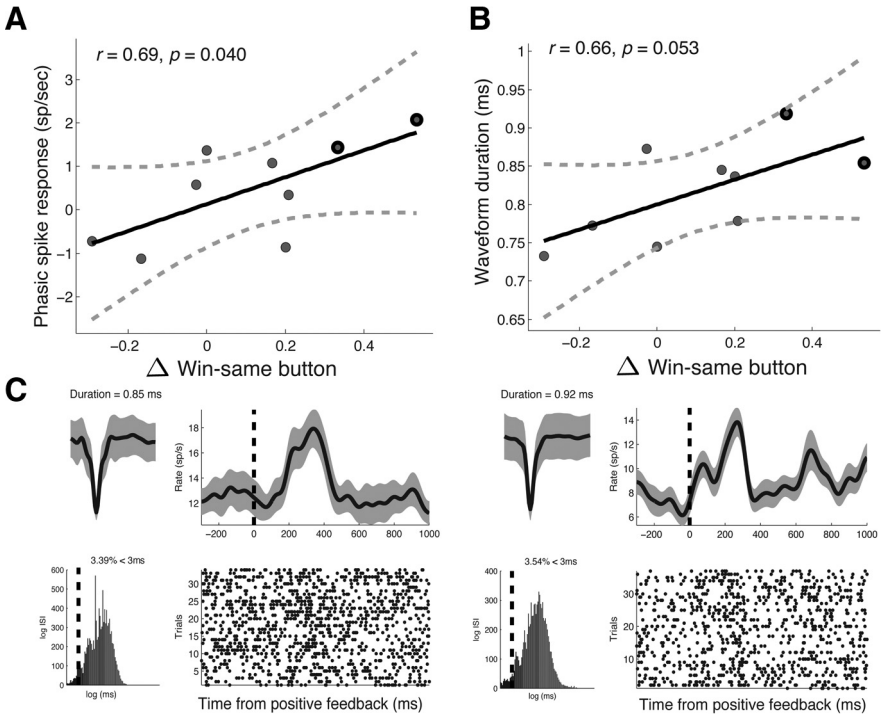

Figure 6. Relation between stimulation-related action bias and recorded neural activity. ***A***, ***B***, Stimulation-related increases in win-same button were positively correlated with postreward phasic responses (***A***) and the mean waveform duration (***B***) of multiunit activity recorded during Stage 1. Each dot represents a subject, the solid black line is the regression slope, and the dashed lines represent 95% confidence intervals. Nine of the 11 subjects contributed to this analysis (we were unable to obtain recordings from Subject #3 and we did not identify spiking activity from Subject #11; see Materials and Methods). ***C***, Example waveforms and postreward phasic responses of unit activity from the two subjects who showed the greatest increases in win-same button (outlined in black in ***A*** and ***B***). For each unit, we show the average waveform (top left, gray shading marks the SD), the interspike interval (bottom left, black line marks 3 ms), the average postreward firing response (top right, smoothed with a Gaussian kernel of half-width = 75 ms; gray shading indicates SE of mean), and the spike raster after reward trials. Dashed black line indicates reward onset.

are sufficient to guide learning (Schultz et al., 1997; Reynolds et al., 2001; Bayer and Glimcher, 2005; Tsai et al., 2009); however, they may not generalize to human learning because animals in these studies have typically undergone long periods of intense training. Conversely, human studies have not demonstrated a functional role for phasic DA activity in learning. Demonstrations of altered learning in patients with PD (Knowlton et al., 1996; Foerde et al., 2013) and in association with pharmacological administration of DA agonists (Frank et al., 2004; Rutledge et al., 2009) may be driven by changes in tonic DA levels throughout the brain (that may alter learning through nonspecific increase in motivation or arousal; Niv et al., 2007). Because SN stimulation has been shown to manipulate local neuronal activity (Histed et al., 2009; Clark et al., 2011), our finding that SN microstimulation during the phasic postreward interval alters learning provides direct evidence for the functional role of phasic SN activity in human reinforcement learning.

Relation to action–reward associations and DA activity

There are several explanations for the observed stimulation-related decrease in learning. One possibility is that microstimulation disrupted the encoding of RPEs or increased the noise in the choice policy, both of which would result in increasingly random choices after stimulation trials. Alternatively, microstimulation may have strengthened competing action–reward associations, which would not result in random item choices, but a bias toward repeating the same button press after reward trials ("win-same button").

We provide the following support for the hypothesis that stimulation enhances action–reward associations. First, we found a positive correlation between stimulation-related decreases in performance and stimulation-related increases in win-same but-

ton. Second, we showed (via simulations of the *Q* learning model) that decreased learning rate or increased noise in the choice policy provide insufficient explanations of stimulation-related changes in behavior. Third, we showed that changes in the relative strength of action–reward associations in a hybrid *AQ* model can capture the major stimulation-related behavioral changes, including the positive correlation between stimulation-related decreases in accuracy and increased win-same button. Finally, we fit the hybrid *AQ* model quantitatively to subjects' choice data and showed that stimulation-related decreases in accuracy were better explained by increases in the relative strength of action–reward associations than decreases in learning rate or increases in decision-making noise. Therefore, SN microstimulation may have disrupted learning during the task by strengthening action–reward rather than stimulus–reward associations.

One might expect strengthened action–reward associations after enhancement of phasic DA activity in the SN. Previous work has shown that SN DA neurons predominantly send their efferent projections to dorsal striatal regions, which mediate action selection (Haber et al., 2000; Lau and Glimcher, 2008); therefore, these neurons are hypothesized to preferentially strengthen action–reward associations during reinforcement learning (Montague et al., 1996; O' Doherty et al., 2004; Frank and Surmeier, 2009). Consistent with this hypothesis, we found that stimulation-related increases in win-same button were most prominent when the microelectrode was positioned near neuronal populations that demonstrated properties characteristic of DA neurons, particularly broad waveforms and phasic postreward responses (Zaghloul et al., 2009; Ungless and Grace, 2012). Because SN DA neurons are coupled via electrical junctions (Vandercasteele et al., 2005), stimulation near a small cluster of DA neurons might result in a spread of depolarization through a larger DA population. This interpretation is in agreement with a previous rodent study showing that microstimulation of certain SN subregions enhances action reinforcement and strengthens corticostriatal synapses in a dopamine-dependent manner (Reynolds et al., 2001).

If SN DA neurons predominantly modulate action–reward associations, then their phasic responses should be more strongly modulated by the reward expectation associated with particular actions, rather than particular stimuli. This has not been tested directly in the human SN—the only previous demonstration of RPE-like responses from human SN DA neurons occurred during a reinforcement learning task with consistent stimulus-response mapping (Zaghloul et al., 2009). In that study, rewards were contingent on particular actions taken by the subjects, leaving open the possibility that SN DA responses were modulated by action-related reward expectancies rather than stimulus-related reward expectancies.

Stimulation during negative feedback

Even though we observed reliable changes in learning performance when SN microstimulation was provided during positive feedback, we were unable to observe such changes when microstimulation was provided during negative feedback. These findings are consistent with previous studies suggesting that the DA system encodes positive RPEs more reliably than negative RPEs (but see Frank et al., 2004; Bayer and Glimcher, 2005, 2007; Rutledge et al., 2009). It is possible that microstimulation manipulated SN-mediated action–reward associations after negative outcomes, but that the SN′s influence on learning was mitigated by the influence of a separate nondopaminergic system that mediates learning from negative outcomes (e.g., serotonin; Daw et al., 2002). In this case, the behavioral changes associated with negative feedback stimulation might be subtle and may become evident with more data. Furthermore, because the effects of negative feedback stimulation were always tested after the effects of positive feedback stimulation, we cannot rule out a potential order effect. Future studies are needed to resolve this potential confound.

Limitations

The interpretation that SN microstimulation strengthened action–reward associations by enhancing DA responses is supported by subjects' behavior after stimulation trials and functional properties of the neural population near the electrode and is consistent with findings from previous studies. However, there are important limitations to consider. First, although we found a positive relation between stimulation-related decreases in performance and increases in win-same button, we were unable to find a reliable increase in win-same button across subjects. It may be the case that SN microstimulation had heterogeneous effects on our subjects—in some subjects, it may have enhanced DA activity and strengthened action–reward associations, whereas in other subjects, it may have disrupted stimulus–reward associations by inhibiting RPE encoding (Tepper et al., 1995; Morita et al., 2012; Pan et al., 2013), possibly by an enhancement of GABA-ergic neurons in the SN, which are known to provide inhibitory inputs onto DA neurons.

Second, it is important to consider the tendency of patients with PD to perseverate during cognitive tasks when interpreting our results (Cools et al., 2001). Rutledge et al. (2009) showed that patients with PD demonstrate choice perseveration during reinforcement learning that is dependent on DA levels but independent of reward history. Because stimulus-response mapping was consistent during their study, the observed perseverative effect may be specific to action selection rather than item choices. Therefore, the stimulation-related increases win-same button that we observed in some of our subjects may also be explained by increased action perseveration. However, because action perseveration is not related to reward history, one would expect to observe a similar behavioral change in association with positive and negative feedback stimulation, which we did not observe.

Finally, the population we studied—patients undergoing DBS surgery for PD—is known to have degeneration of DA neurons in SN. Ideally, one would like to characterize the behavioral changes associated with SN microstimulation in healthy human subjects, but at present, SN microstimulation may not be ethically conducted in any other human population. Certainly, this poses a challenge for interpreting findings concerning the functional role of SN neurons in patients who have degenerative disease. However, histological studies in PD patients (Damier et al., 1999) and electrophysiological studies in rat models of PD (Hollerman and Grace, 1990; Zigmond et al., 1990) and in humans (Zaghloul et al., 2009) indicate that a significant population of viable DA neurons remain in the parkinsonian SN. By demonstrating altered reinforcement learning performance in association with SN microstimulation, our results suggest that these remaining neural processes may be functionally relevant for choice behavior.

Conclusions

In this study, we show that manipulation of phasic SN activity via electrical microstimulation during rewards disrupted performance on a reinforcement learning task in which rewards were contingent on stimuli rather than actions. The greatest decreases in learning were observed when subjects showed an increased

propensity to repeat the same action after rewards, suggesting that SN microstimulation strengthened action–reward associations rather than stimulus–reward associations during the task. Although future studies are needed to rule out alternative explanations for the observed results, such as disrupted RPE encoding or increased action perseveration, our findings provide support for the hypothesis that SN DA neurons preferentially strengthen action–reward associations during reinforcement learning.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716–723. CrossRef

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 dopaminergic neuron spike trains in the awake primate. J Neurophysiol 98:1428–1439. CrossRef Medline

Clark KL, Armstrong KM, Moore T (2011) Probing neural circuitry and function with electrical microstimulation. Proc Biol Sci 278:1121–1130. CrossRef Medline

Cools R, Barker RA, Sahakian BJ, Robbins TW (2001) Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. Cereb Cortex 11:1136–1143. CrossRef Medline

Damier P, Hirsch EC, Agid Y, Graybiel AM (1999) The substantia nigra of the human brain. II. Patterns of loss of dopamine-containing neurons in Parkinson's disease. Brain 122:1437–1448. CrossRef Medline

Daw ND, Kakade S, Dayan P (2002) Opponent interactions between serotonin and dopamine. Neural Netw 15:603–616. CrossRef Medline

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876–879. CrossRef Medline

Foerde K, Race E, Verfaellie M, Shohamy D (2013) A role for the medial temporal lobe in feedback-driven learning: evidence from amnesia. J Neurosci 33:5698–5704. CrossRef Medline

Frank MJ, Surmeier DJ (2009) Do substantia nigra dopaminergic neurons differentiate between reward and punishment? J Mol Cell Biol 1:15–16. CrossRef Medline

Frank MJ, Seeberger LC, O'reilly RC (2004) By carrot or by stick: Cognitive reinforcement learning in parkinsonism. Science 306:1940–1943. CrossRef Medline

Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007) Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl Acad Sci U S A 104:16311–16316. CrossRef 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

Grattan L, Rutledge R, Glimcher P (2011) Program No. 732.12. 2011 Neuroscience Meeting Planner, San Diego, CA: Society for Neuroscience. Online.

Haber SN, Fudge JL, McFarland NR (2000) Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. J Neurosci 20:2369–2382. Medline

Henny P, Brown M, Northrop A, Faunes M, Ungless MA, Magill PJ, Bolam JP (2012) Structural correlates of heterogeneous in vivo activity of midbrain dopaminergic neurons. Nat Neurosci 15:613–619. CrossRef Medline

Histed MH, Bonin V, Reid RC (2009) Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation. Neuron 63:508–522. CrossRef Medline

Hollerman JR, Grace AA (1990) The effects of dopamine-depleting brain lesions on the electrophysiological activity of rat substantia nigra dopamine neurons. Brain Res 533:203–212. CrossRef Medline

Jaggi JL, Umemura A, Hurtig HI, Siderowf AD, Colcher A, Stern MB, Baltuch GH (2004) Bilateral subthalamic stimulation of the subthalamic nucleus in Parkinson's disease: surgical efficacy and prediction of outcome. Stereotact Funct Neurosurg 82:104–114. CrossRef Medline

Knowlton BJ, Mangles JA, Squire LR (1996) A neostriatal habit learning system in humans. Science 273:1399–1402. CrossRef Medline

Lafreniere-Roula M, Hutchison WD, Lozano AM, Hodaie M, Dostrovsky JO (2009) Microstimulation-induced inhibition as a tool to aid targeting the ventral border of the subthalamic nucleus. J Neurosurg 111:724–728. CrossRef Medline

Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. Neuron 58:451–463. CrossRef Medline

Ma SY, Rinne JO, Collan Y, Röyttä M, Rinne UK (1996) A quantitative morphometrical study of neuron degeneration in the substantia nigra in Parkinson's disease. J Neurol Sci 140:40–45. CrossRef Medline

Matsumoto M, Hikosaka O (2009) Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature 459:837–841. CrossRef 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

Morita K, Morishima M, Sakai K, Kawaguchi Y (2012) Reinforcement learning: computing the temporal difference of values via distinct corticostriatal pathways. Trends Neurosci 35:457–467. CrossRef Medline

Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology 191:507–520. CrossRef Medline

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304:452–454. CrossRef Medline

Pan WX, Brown J, Dudman JT (2013) Neural signals of extinction in the inhibitory microcircuit of the ventral midbrain. Nat Neurosci 16:71–78. CrossRef Medline

Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. Nature 435:1102–1107. CrossRef Medline

Reynolds JN, Hyland BI, Wickens JR (2001) A cellular mechanism of reward-related learning. Nature 413:67–70. CrossRef Medline

Rossion B, Pourtois G (2004) Revisiting Snodgrass and Vanderwart's object set: The role of surface detail in basic-level object recognition. Perception 33:217–236. CrossRef Medline

Rutledge RB, Lazzaro SC, Lau B, Myers CE, Gluck MA, Glimcher PW (2009) Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. J Neurosci 29:15104–15114. CrossRef Medline

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

Shiner T, Seymour B, Wunderlich K, Hill C, Bhatia KP, Dayan P, Dolan RJ (2012) Dopamine and performance in a reinforcement learning task: evidence from Parkinson's disease. Brain 135:1871–1883. CrossRef Medline

Sugrue LP, Corrado GS, Newsome WT (2005) Choosing the greater of two goods: neural currencies for valuation and decision making. Nat Rev Neurosci 6:363–375. CrossRef Medline

Sutton, R. and Barto, A (1990) Time-derivative models of Pavlovian reinforcement. In: Learning and computational neuroscience: foundations of adaptive networks (Gabriel M, Moore J, eds), pp 497–537. Cambridge, MA: MIT.

Tepper JM, Martin LP, Anderson DR (1995) GABA-A receptor-mediated inhibition of rat substantia nigra dopaminergic neurons by pars reticulata projection neurons. J Neurosci 15:3092–3103. Medline

Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K (2009) Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science 324:1080–1084. CrossRef Medline

Ungless MA, Grace AA (2012) Are you or aren't you? Challenges associated with physiologically identifying dopamine neurons. Trends Neurosci 35:422–430. CrossRef Medline

Vandecasteele M, Glowinski J, Venance L (2005) Electrical synapses between dopaminergic neurons of the substantia nigra pars compacta. J Neurosci 25:291–298. CrossRef Medline

Zaghloul KA, Blanco JA, Weidemann CT, McGill K, Jaggi JL, Baltuch GH, Kahana MJ (2009) Human substantia nigra neurons encode unexpected financial rewards. Science 323:1496–1499. CrossRef Medline

Zigmond MJ, Abercrombie ED, Berger TW, Grace AA, Stricker EM (1990) Compensations after lesions of central dopaminergic neurons: some clinical and basic implications. Trends Neurosci 13:290–296. CrossRef Medline