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, [3]Department of Neurosurgery, Perelman School of Medicine
, University of Pennsylvania, Philadelphia, PA 19103
[4] Drexel University School of Biomedical Engineering, Science and Health Systems, Philadelphia, PA 19103
*These authors contributed equally to this work.

Correspondence should be addressed to either:
G.H.B (baltuchg@mail.med.upenn.edu)
Department of Neurosurgery
Perelman School of Medicine
235 South 8th Street
Philadelphia, PA 19106
or,
M.J.K. (kahana@psych.upenn.edu).
Department of Psychology
University of Pennsylvania
3401 Walnut St., Room 303C
Philadelphia, PA 19104

Abbreviated title: SN microstimulation alters human learning

Word Count: 6,814 Abstract: 153; Introduction: 497; Materials and Methods: 1,735; Results: 1,133; Discussion 1,485; References 1,233; Legends 578

Abstract

Animal studies have shown that phasic bursts of substantia nigra dopaminergic (DA) neurons strengthen action-reward associations during reinforcement learning (Reynolds et al., 2001) but their role in human learning is not known. Here we applied microstimulation in the substantia nigra (SN) of 11 patients undergoing deep brain stimulation (DBS) surgery for the treatment of Parkinson's disease (PD) as they performed a two-alternative probability learning task, where rewards were contingent on stimuli, rather than actions. Subjects demonstrated decreased learning following reward trials that were accompanied by phasic SN microstimulation, compared to reward trials without stimulation. These decreases in learning were associated with an increased bias towards repeating actions following stimulation trials, suggesting that stimulation may have decreased learning by strengthening action-reward associations, rather than stimulus-reward associations. Our findings build on previous studies that implicate SN DA neurons in preferentially strengthening action-reward associations during reinforcement learning (Montague et al., 1996; Haber et al., 2000).

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. Electrophysiological recordings in the animal have shown that dopaminergic (DA) neurons in the ventral tegmental area 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 cortico-striatal 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 to 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 non-specific 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 previously applied in association with a cognitive task. Here we applied microstimulation during the 500-ms following a subset of feedback trials as subjects performed a reinforcement learning task, where rewards were contingent on stimuli, rather than actions (putative DA neurons in the human SN have been shown to display RPE-like responses during this post-feedback time interval; Zaghloul et al. (2009)). If phasic SN responses preferentially strengthen action-reward associations during reinforcement learning,

stimulation following reward trials should induce a bias to repeating actions, rather than stimuli, and disrupt learning during the task.

Materials and Methods

Participants: Eleven patients undergoing deep brain stimulation (DBS) surgery for the treatment of Parkinson's Disease volunteered to take part in this study (8 male, 3 female, age = 63 ± 7, mean ± S.D). Subjects provided their informed consent during pre-operative 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); hence 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.

Intra-operative methods: During surgery, intra-operative microelectrode recordings (obtained from a 1 μm diameter tungsten tip electrode advanced with a power-assisted microdrive) were used to identify the substantia nigra (SN) and the subthalamic nucleus (STN) as per routine clinical protocol (Jaggi et al., 2004). 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 IRB. 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 participant.

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 (Figure 1b; Sugrue et al. (2005)). Subjects were asked to choose between pairs of items and probabilistically received positive or negative feedback following each choice (Figure 1a). 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. In order 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, participant choice, and feedback presentation. In the event of positive feedback ("wins") the screen turned green and the audible ring of a cash register was presented. In the event of negative feedback ("losses") the screen turned red and an audible buzz 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. To ensure that reward probabilities observed by the participant did not drastically fluctuate over the course of the session, we simulated reward probabilities during each session deterministically, rather than pseudorandomly. For instance, to simulate a reward probability of 0.8, we ensured that 4 out of every 5 selections of that stimulus result in positive feedback.

Each session consisted of 150 trials ($\approx$ 15 minutes of testing time) and was subdivided into three stages (50 trials each, Figure 1b,c). Each stage consisted of two novel pairs of stimuli which were matched in relative reward rate. The relative reward rates for each pair were set to 0.8 vs. 0.2. If the participant selected the high-probability item on at least 80% of trials on stage 1, the relative reward rates for pairs in subsequent stages were set to 0.7 vs. 0.3 to encourage learning during the remainder of the session, otherwise, they remained the same. The two item pairs within each stage were presented in alternating trains of 3 to 6 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 following a subset of feedback trials (see *Stimulation parameters*). During stage 2, one of the pairs was

associated with SN microstimulation during *positive feedback* following 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* following an 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.

When possible, subjects first performed the task during preoperative consultation, but in all cases, the task was reviewed with subjects during the morning of surgery. Further instructions were provided prior to beginning the task intra-operatively. Subject #3 did not perform stage 1 due to a technical difficulty during the experiment, but completed stages 2 and 3 of the task (Table 1). 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 following feedback presentation during the learning task using an FHC Pulsar 6b microstimulator using the following parameters: bi-phasic, 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 non-human primate VTA (Grattan et al., 2011). Stimulation trials were not signaled to subjects in any other manner; none of the subjects reported a perceptual change following application of microstimulation.

Reinforcement learning model simulations: To study the correlations between various behavioral measures in our task, we simulated the performance of a standard reinforcement learning model ($Q-$learning; Sutton and Barto (1990)) on a two-alternative probability learning task with inconsistent stimulus-response mapping. Each 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. The Q model maintains independent estimates of reward expectation (Q) values for each option i at each time t. A choice is probabilistically generated on each trial by comparing the Q values of available options on that trial using the following logistic function. $P_i(t) = \frac{\exp(Q_i(t)/\beta)}{\sum_j \exp(Q_j(t)/\beta)}$. 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)). It was set to a value of 0.2 for all our simulations. Once the model probabilistically selected an item on a trial, we assumed that there was no additional noise in mapping the stimulus to the associated action on a given trial. 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) incorporate reinforcements from many previous trials. We simulated the performance of 34 $Q-$model agents that varied in their α values (0.01 to 1, with a step size of 0.03; Frank et al. (2007)). We simulated the performance of these agents on 1000 randomly generated sessions.

Extracting spiking activity from microelectrode recordings: We obtained microelectrode recordings as subjects performed stage 1 prior to applying microstimulation during the experiment. Because these recordings were of a relatively short duration ($\approx$ 5 min.) and 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 multi-unit 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

(≈3.5 S.D 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 multi-unit. From each multi-unit, we extracted two features that are characteristic of DA activity — the mean waveform duration and the phasic post-reward 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 post-reward response as the difference between the average spike rate during 0-500 ms post-reward interval, and that during the -250-0 and 500-750 ms intervals. We did not consider responses following negative outcomes because dopaminergic neurons are not homogenous in their responses following negative outcomes (Matsumoto and Hikosaka, 2009). We obtained multi-unit activity from 9 of the 11 subjects. We were unable to obtain recordings from one subject (#3) and could not distinguish spiking activity from noise contamination in another subject (#11).

Results

Eleven subjects performed a two–alternative probability learning task where they selected between pairs of items (images of common objects) and probabilistically

received abstract rewards ("wins") or punishments ("losses") following each choice (Figure 1a). Subjects were instructed that one item in each pair carried a higher reward probability that the other item in the pair, and that their goal was to maximize the number of rewards 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 action-reward associations in order 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*). During stage 1, we did not provide stimulation in association with either item pair (SHAM) so that participants 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 towards 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, Figure 1b,c). By comparing learning on the STIM and SHAM pair within each stage, we sought to assess the effects of SN microstimulation on learning.

[Figure 1 about here.]

During stage 2, we assessed the effect of stimulation on reward learning by applying stimulation following 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 to the SHAM pair during this stage ($t(10) = 2.56$, $p = 0.029$, Figure 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 following stimulation trials during this stage, we compared subjects' tendencies to repeat their selection of the high-reward probability item following rewards ("win-stay") on the STIM+ and the SHAM pair. We found that subjects reliably demonstrated decreased win-stay following reward trials accompanied by stimulation compared reward trials without stimulation ($t(10) = 2.71, p = 0.022$). Thus, subjects demonstrated decreased learning following reward trials that were accompanied by phasic SN microstimulation compared to reward trials without stimulation. During stage 3, we applied stimulation following negative feedback associated with the low-reward probability item on one item pair ($STIM^-$) to study the influence of 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 (Figure 2) or their probability repeating an item choice following stimulation trials ($p's > 0.3$).

[Figure 2 about here.]

Our main finding is that SN microstimulation following 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 towards repeating actions rather than stimuli following 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 following 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, subjects should show an increased bias towards repeating the same button following high-probability reward trials ("win-same button") on the STIM^{+} pair compared to 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 ($r = 0.77, p = 0.006$, Figure 3a.). Thus, subjects who showed the greatest stimulation-related decreases in learning also showed an increased bias towards repeating actions following stimulation trials.

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 a decrease in accuracy is necessarily associated with an increased win-same button during our task. To assess whether this was the case, we simulated the performance of a standard reinforcement learning model performing a two-alternative probability learning task with inconsistent stimulus-response mapping (*Materials and Methods*, Figure 3b). We found that decreases in the learning rate of the model resulted in decreases in accuracy and win-stay, but no accompanying change in win-same button. Thus, the positive relation between decreased accuracy and win-same button is not a necessary result of the task design.

[Figure 3 about here.]

These results suggest that stimulation may have strengthened action-reward associations during the task, that may be related to enhanced phasic DA activity in the SN (Reynolds et al., 2001; Montague et al., 1996). 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. Thus, 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 multi-unit spiking activity from each recording and extracted two features that are characteristic of DA activity—average waveform duration and the phasic post-reward response (see

Materials and Methods; Ungless and Grace (2012); Zaghloul et al. (2009)). We found positive correlations between stimulation-related increases in win-same button and both the phasic post-reward response (Figure 4a, Pearson's $r = 0.69, p = 0.040$) and the mean waveform duration of recorded multi-unit activity (Figure 4b, Pearson's $r = 0.66$, $p = 0.053$). Multi-units recorded from the two subjects that showed the greatest increases in win-same button showed broad waveforms (0.85 ms, and 0.92 ms) and phasic post-reward bursts that were visible in the spike raster (+2.07 spikes/sec, and +1.43 spikes/sec; Figure 4c). 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.

[Figure 4 about here.]

[Table 1 about here.]

[Table 1 about here.]

Discussion

We applied electrical microstimulation in substantia nigra (SN) of 11 patients undergoing deep brain stimulation (DBS) surgery for the treatment of Parkinson's disease (PD) as they performed a two-alternative probability learning task, where rewards were contingent on stimuli rather than actions. Subjects were required to strengthen

stimulus-reward associations, rather than action-reward associations in order to perform well on the task. We found that SN microstimulation applied following reward trials disrupted learning compared to a control learning condition.

Phasic SN activity is functionally important for human reinforcement learning. By showing the SN microstimulation during the phasic post-reward interval alters performance during the task, our findings provide an important bridge between animal and human studies of learning. Animal studies have shown that the phasic activity of DA neurons signal positive reward prediction errors (RPEs; Schultz et al. (1997); Bayer and Glimcher (2005)) that are sufficient to guide learning (Reynolds et al., 2001; Tsai et al., 2009) however, several factors limit the generalizability of these studies to human behavior. For example, animals in these studies have typically undergone long periods of intense training, whereas much of human learning occurs in novel situations. On the other hand, human studies of learning have not demonstrated a functional role for phasic DA activity in learning. Studies have shown that reinforcement learning performance is altered in patients with PD (Knowlton et al., 1996; Foerde et al., 2013) who have degeneration of SN DA neurons, and in association with pharmacological administration of DA agonists (Frank et al., 2004; Rutledge et al., 2009). However, both PD and DA agonists manipulate tonic DA levels throughout the brain which might alter performance on learning tasks through non-specific 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 post-reward 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, which would result increasingly random choices following stimulation trials. Alternatively, microstimulation may have strengthened competing reward-action associations, which would result in random item choices, but a bias towards repeating the same button press following reward trials ("win-same button"). We found a positive correlation between stimulation-related decreases in performance and stimulation-related increase in win-same button, a pattern that would not be expected following a simple decrease in stimulus-reward learning (as shown by our model simulations). Thus, 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 following 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); thus 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 post-reward 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 cortico-striatal 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 directly tested in the human SN—the only previous demonstration of RPE-like responses from human SN DA neurons occurred during a reinforcement learning task 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. Functional neuroimaging studies have shown

evidence for DA-dependent RPE encoding in the human ventral striatum during a learning task with inconsistent stimulus-response mapping (Pessiglione et al., 2006) however, these changes may have been driven by DA populations in the ventral tegmental area which project strongly to the ventral striatum (Haber et al., 2000) and may preferentially strengthen stimulus-reward associations.

Stimulation following negative feedback. Even though we observed reliable changes in learning performance when SN microstimulation was provided following positive feedback, we were unable to observe such changes when microstimulation was provided following negative feedback. These findings are consistent with previous studies which suggesting that the DA system exclusively encode positive RPEs (Bayer and Glimcher (2005); Rutledge et al. (2009); although, see Frank et al. (2004)). It is possible that microstimulation manipulated SN-mediated reward-action associations following negative outcomes, but that the SN's influence on learning was mitigated by the influence of separate non-dopaminergic system that mediates learning from negative outcomes (e.g., serotonin; Daw et al. (2002)). Then, the behavioral changes following negative feedback stimulation might be subtle and may become evident with more data.

Limitations The interpretation that SN microstimulation strengthened reward-action associations by enhancing DA responses is supported by subjects' behavior following stimulation trials, 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 reward-action associations, whereas in other subjects it may disrupted reward-stimulus associations by inhibiting RPE encoding (possibly by an enhancement of GABA-ergic neurons in the SN, which are known to provide inhibitory inputs onto DA neurons; Tepper et al. (1995); Morita et al. (2012); Pan et al. (2013)). If the greatest decreases in performance were related to enhanced reward-action associations, one would observe a positive relation between decreased learning and win same-button, but no reliable increase in win same-button.

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, which was dependent on DA levels, but independent of reward history. Because stimulus-response mapping was consistent during this study, this perseverative effect may be specific to action selection rather than item choices. Thus, 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 following 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 the challenge of 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; Wang et al., 2010), and 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 following rewards disrupted performance on a reinforcement learning task where 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 following rewards, suggesting that SN microstimulation strengthened action-reward associations, rather than stimulus-reward

associations during the task. Our findings are consistent with previous work implicating SN DA neurons in preferentially strengthening action-reward associations during reinforcement learning (Reynolds et al., 2001). However, future studies are needed to rule out alternative explanations for the observed results such as disrupted RPE-encoding or increased action perseveration.

References

Bayer, H. and Glimcher, P. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 47:129–141.

Clark, K., Armstrong, K., and Moore, T. (2011). Probing neural circuitry and function with electrical microstimulation. *Proceedings of the Royal Society B: Biological Sciences*.

Cools, R., Barker, R., Sahakian, B., and Robbins, T. (2001). Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. *Cerebral Cortex*, 11:1136–1143.

Damier, P., Hirsch, E., Agid, Y., and Graybiel, A. M. (1999). The Substantia Nigra of the human brain ii. patterns of loss of dopamine-containing neurons in Parkinson's disease. *Brain*, 122:1437–1448.

Daw, N., Kakade, S., and Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks*, 15(4-6):603–616.

Daw, N., O'Doherty, J., Dayan, P., Seymour, B., and Dolan, R. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441:876–879.

Foerde, K., Race, E., Verfaellie, M., and Shohamy, D. (2013). A role for the medial temporal lobe in feedback-driven learning: Evidence from amnesia. *Journal of Neuroscience*, 33(13):5698–5704.

Frank, M., Moustafa, A., Haughey, H., Curran, T., and Hutchison, K. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences, USA*, 104:16311–16316.

Frank, M. and Surmeier, D. (2009). Do substantia nigra dopaminergic neurons differentiate between reward and punishment? *Journal of Molecular Cell Biology*, 1:15–16.

Frank, M. J., Seeberger, L. C., and O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science*, 306:1940–1943.

Glimcher, P. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences, USA*, 108(3):15647–15654.

Grattan, L., Rutledge, R., and Glimcher, P. (2011). Increased dopamine concentrations increase the perceived value of an action. In *Program No. 732.12. Society for Neuroscience Meeting Planner*, San Diego, CA. Society for Neuroscience.

Haber, S. N., Fudge, J. L., and McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *Journal of Neuroscience*, 20(6):2369–2382.

Henny, P., Brown, M., Northrop, A., Faunes, M., Ungless, M., Magill, P., and Bolam, J.

(2012). Structural correlates of heterogeneous in vivo activity of midbrain dopaminergic neurons. *Nature Neuroscience*, 15(4):613–619.

Histed, M., Bonin, V., and Reid, C. (2009). Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation. *Neuron*, 63:508–522.

Hollerman, J. and Grace, A. (1990). The effects of dopamine-depleting brain lesions on the electrophysiological activity of rat Substantia Nigra dopamine neurons. *Brain Research*, 533:203–212.

Jaggi, J., Umemura, A., Hurtig, H., Siderowf, A., Colcher, A., Stern, M., and Baltuch, G. (2004). Bilateral subthalamic stimulation of the subthalamic nucleus in Parkinson's disease: surgical efficacy and prediction of outcome. *Stereotactc & Functional Neurosurgery*, 82:104–14.

Knowlton, B., Mangles, J., and Squire, L. (1996). A neostriatal habit learning system in humans. *Science*, 273(5280):1399–1402.

Lafreniere-Roula, M., Hutchinson, W., Lozano, A., Hodaie, M., and Dostrovsky, J. (2009). Microstimulation-induced inhibition as a tool to aid targeting the ventral border of the subthalamic nucleus. *Journal of Neurosurgery*, 111(4):724–728.

Lau, B. and Glimcher, P. (2008). Value representations in the primate striatum during matching behavior. *Neuron*, 58(3):451–463.

Ma, S., Rinne, J., Collan, Y., Roytta, M., and Rinne, U. (1996). A quantitative morphometrical study of neuron degeneration in the substantia nigra in Parkinson's disease. *Journal of the neurological sciences*, 140(1-2):40–45.

Matsumoto, M. and Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, 459(11):837–841.

Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of Neuroscience*, 16:1936–1947.

Morita, K., Morishima, M., Sakai, K., and Kawaguchi, Y. (2012). Reinforcement learning: computing the temporal difference of values via distinct corticostriatal pathways. *Trends in Neurosciences*, 35(8):457–467.

Niv, Y., Daw, N., Joel, D., and Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, 191(3):507–520.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304(5669):452–454.

Pan, W. X., Brown, J., and Dudman, J. (2013). Neural signals of extinction in the inhibitory microcircuit of the ventral midbrain. *Nature Neuroscience*, 16(1):71–78.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., and Frith, C. (2006). Dopamine-dependent prediction errors underpin reward-seeking behavior in humans. *Nature*, 442:1042–1045.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(23):1102–1107.

Reynolds, J., Hyland, B., and Wickens, J. (2001). A cellular mechanism of reward-related learning. *Nature*, 413:67–70.

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

Rutledge, R., Lazzaro, S., Lau, B., Myers, C. E., Gluck, M. A., and Glimcher, P. (2009). Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *Journal of Neuroscience*, 29(48):15104–15114.

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

Shiner, T., Seymour, B., Wunderlich, K., Hill, C., Bhatia, K.P., D. P., and Dolan, R. J. (2012). Dopamine and performance in a reinforcement learning task: evidence from parkinson's disease. *Brain*, 135:1871–1883.

Sugrue, L., Corrado, G., and Newsome, W. (2005). Choosing the greater of two goods:

neural currencies for valuation and decision making. *Nature Reviews Neuroscience*, 6:363–375.

Sutton, R. and Barto, A. (1990). Time-derivative models of pavolovian reinforcement. In Gabriel, M. and Moore, J., editors, *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, pages 497–537. MIT Press, Cambridge, MA.

Tepper, J., Martin, L., and Anderson, D. (1995). GABA-A receptor-mediated inhibition of rat Substantia Nigra dopaminergic neurons by pars reticulata projection neurons. *Journal of Neuroscience*, 15(4):3092–3103.

Tsai, H., Zhang, F., Adamatidis, A., Stuber, Garret, S., Bonci, A., Lecea, L., and Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science*, 324(5930):1080–1084.

Ungless, M. and Grace, A. (2012). Are you or aren't you? Challenges associated with physiologically identifying dopamine neurons. *Trends in Neurosciences*, 35:422–30.

Vandercasteele, M., Glowinski, J., and Venance, L. (2005). Electrical synapses between dopaminergic neurons of the substantia nigra pars compacta. *Journal of Neuroscience*, 25(2):291–298.

Wang, Y., Zhang, Q., Ali, U., Gui, Z., Hui, Y., Chen, L., and Wang, T. (2010). Changes in firing rate and pattern of GABA-ergic neurons in subregions of the Substantia Nigra pars reticulata in rat models of Parkinson's Disease. *Brain Research*, 1324:54–63.

Zaghloul, K. A., Blanco, J. A., Weidemann, C. T., McGill, K., Jaggi, J. L., Baltuch, G. H., and Kahana, M. J. (2009). Human Substantia Nigra neurons encode unexpected financial rewards. *Science*, 323:1496–1499.

Zigmond, M., Abercrombie, E., Berger, T. W., Grace, A., and Stricker, E. (1990). Compensations after lesions of central dopaminergic neurons: some clinical and basic implications. *Trends in Neurosciences*, 13:290–296.

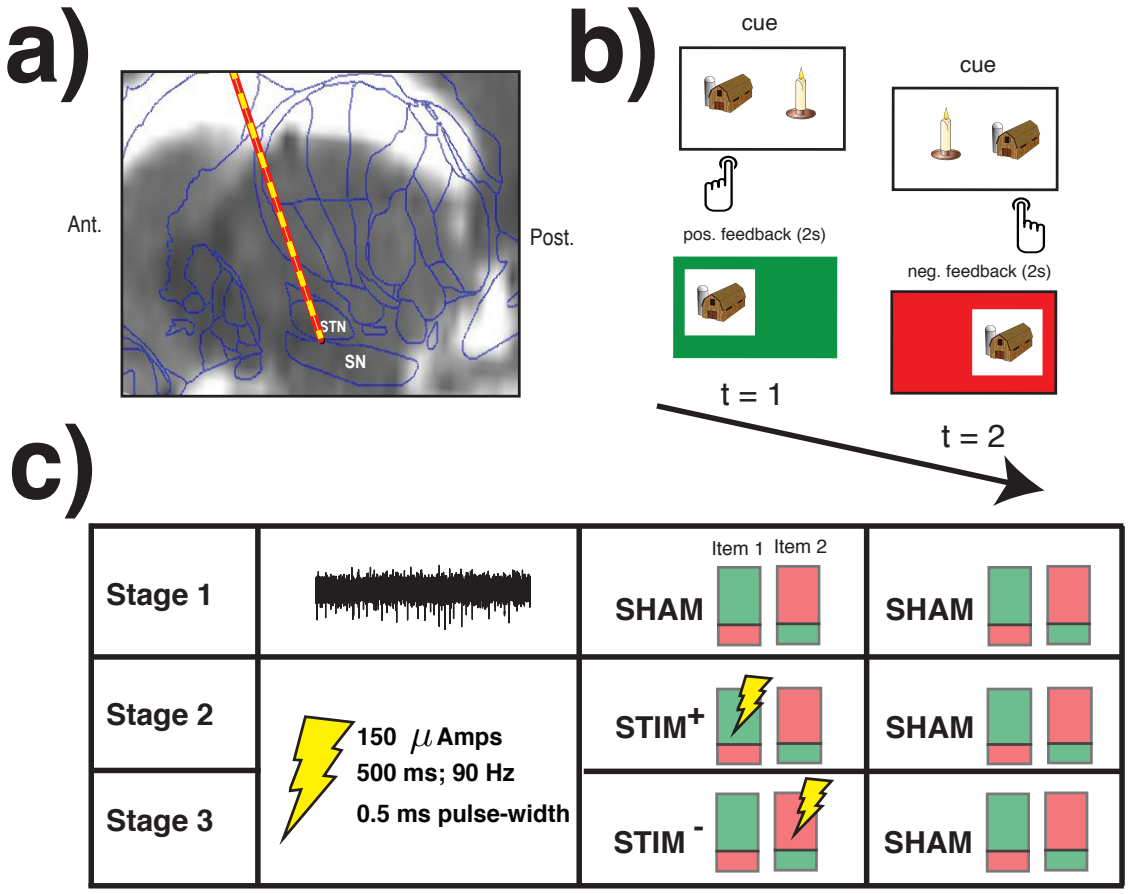

Figure 1: **A.Intra-operative targeting of substantia nigra.** During deep brain stimulation (DBS) surgery, a microelectrode in advanced into the substantia nigra (SN) to map the ventral border of the subthalamic nucleus (STN). An example pre-operative MRI scan (sagittal view) overlaid with a standard 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). During stage 1, we obtained microelectrode recordings from the SN. An example 500 ms high-pass (> 300 Hz) filtered 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*).

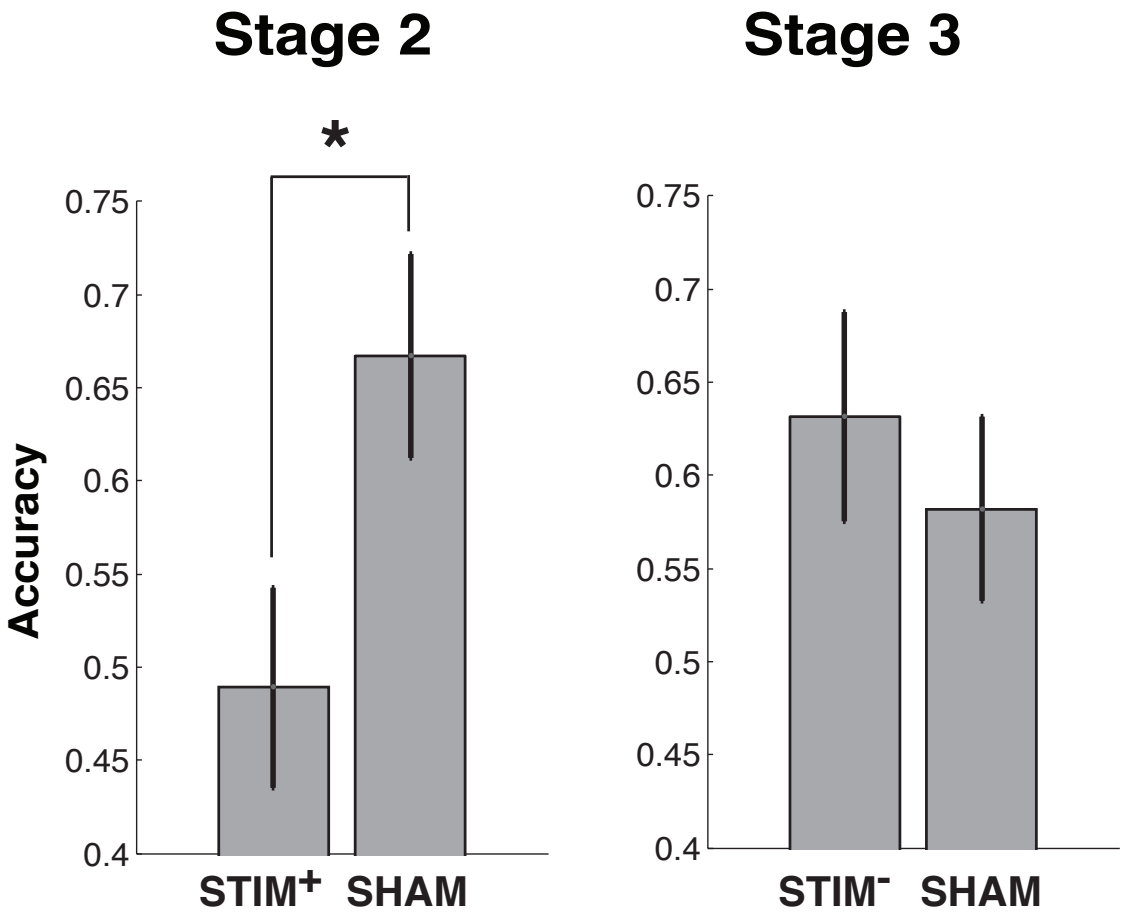

Figure 2: **Effects of stimulation on learning**. To index learning performance on a particular item pair, we computed the probability that subjects chose the item that was associated with a high reward-probability ("accuracy"). During stage 2, subjects demonstrated lower accuracy on the STIM^{+} pair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

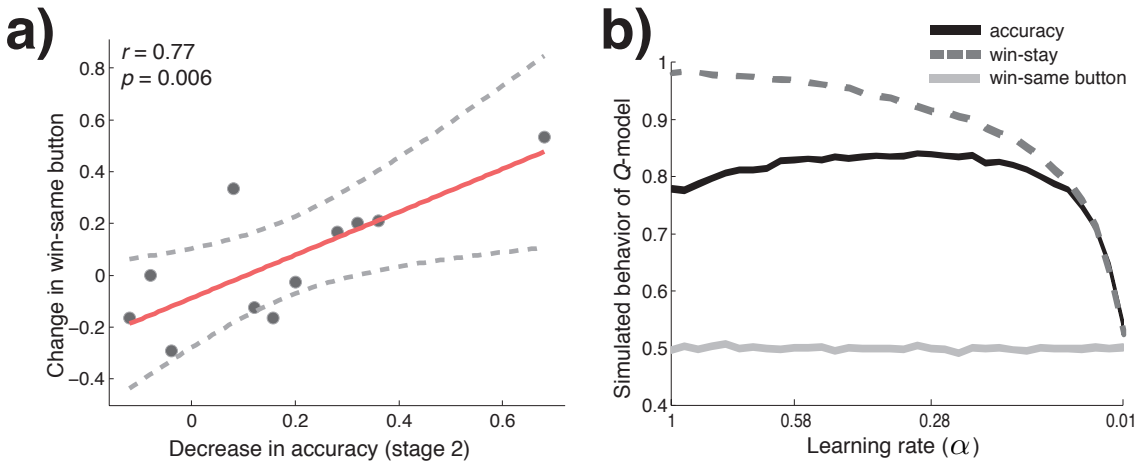

Figure 3: **Relation between decreases in learning and action bias A.** Stimulation-related decreases in accuracy were positively correlated with an increased bias towards repeating a button press following reward trials (win-same button; Pearson's $r = 0.77$, $p = 0.006$). Each dot represents a subject, the solid red line is the regression slope, and the dashed lines represent 95 % confidence intervals. **B.** Simulated behavior of a reinforcement learning algorithm (Q-model) on a two-alternative probability learning task with inconsistent stimulus-response mapping. Accuracy (black line), probability of repeating rewarded items (win-stay, dashed grey line) and probability of repeating rewarded actions (light grey, win-same button) are shown for various learning rates. Decreases in learning rate were accompanied by a decrease in accuracy and a decrease in win-stay, but no change in win-same button.

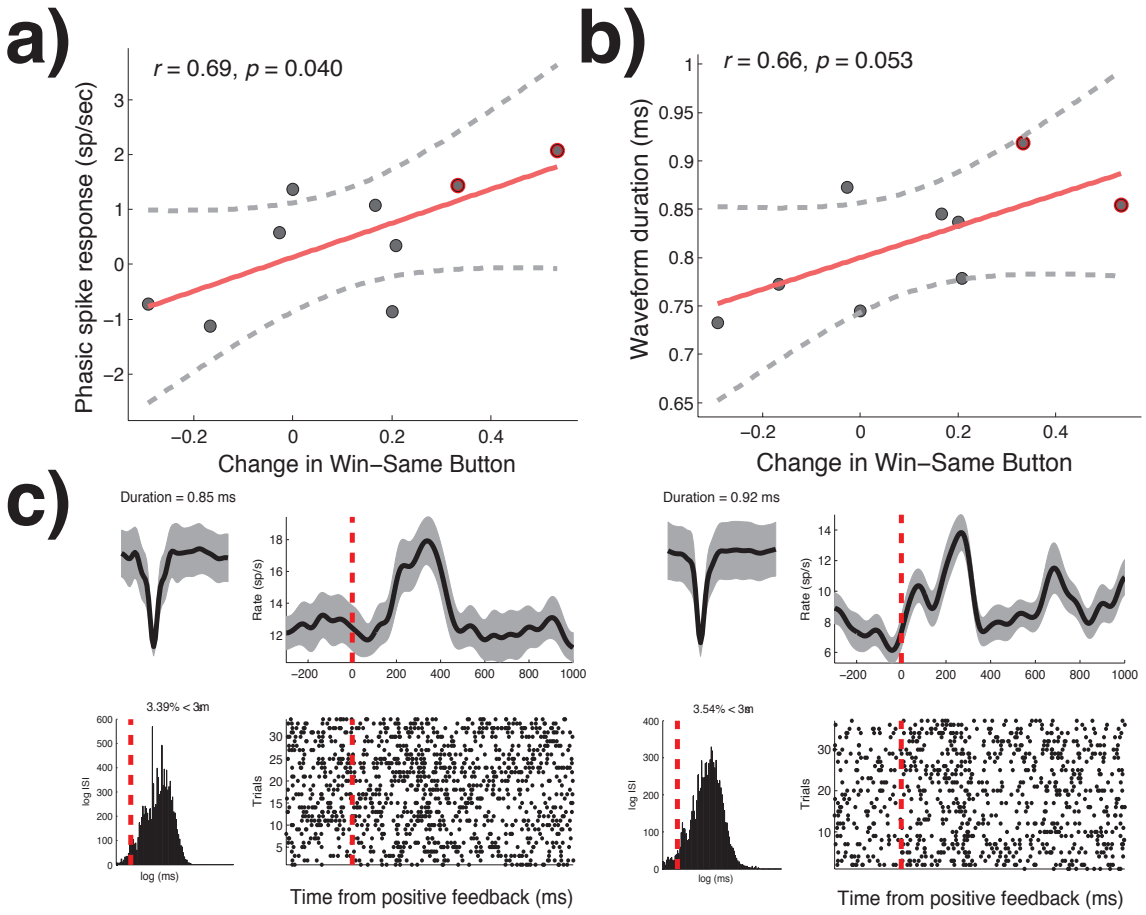

Figure 4: **Relation between stimulation-related action bias and recorded neural activity** Stimulation-related increases in win-same button were positively correlated with post-reward phasic responses (**A.**) and the mean waveform duration (**B.**) of multi-unit activity recorded during stage 1. Each dot represents a subject, the solid red line is the regression slope, and the dashed lines represent 95 % confidence intervals. 9 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 post-reward phasic responses of unit activity from the two subjects who showed the greatest increases in win-same button (outlined in red in panels A and B). For each unit, we show the average waveform (top left, gray shading marks the standard deviation), the inter-spike interval (bottom left, red line marks 3 ms), the average post-reward firing response (top right, smoothed with a Gaussian kernel of half-width=75 ms; gray shading indicates standard error of mean), and the spike raster following reward trials. Dashed red line indicates reward onset.

Subject	Age	Gender	△ accuracy	△ win-stay	△ win-same button	waveform duration	phasic spike response (sp/sec)
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	−0.50	+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	–	–

Table 1: **Summary of pariticpant data**. Columns 4-6 describe behavioral changes during stage 2. Columns 7-8 describe properties of multi-unit activity recorded during stage 1. "-" indicates missing data. We were unable to obtain recordings from subject 3 and did not identify spiking activity from subject 11.