Cognitive control and binding in context-based decision-making : normal and dopamine deviant populations

Wouwe, N.C. van

Chapter 6

Deep brain stimulation of the subthalamic nucleus improves reward-based decision-learning in Parkinson's disease

Abstract

Recently, the subthalamic nucleus (STN) has been shown to be critically involved in decision-making, action selection, and motor control. Here we investigate the effect of deep brain stimulation (DBS) of the STN on reward-based decision-learning in patients diagnosed with Parkinson's disease (PD). We determined computational measures of outcome evaluation and reward prediction from PD patients who performed a probabilistic reward-based decision-learning task with their DBS device turned OFF versus ON. These measures have previously been shown to covary with activation in the nucleus caudatus (outcome evaluation during the early phases of learning) and the putamen (reward prediction during later phases of learning). Stimulation of the STN's motor regions in PD patients served to improve reward-based decision-learning through its effect on activity in frontostriatal motor loops (prominently involving the putamen and, hence, reward prediction). In relatively young patients with relatively short disease duration, the effects of DBS appeared to spread to more cognitive regions of the STN, benefitting loops that connect the caudate to various prefrontal areas important for outcome evaluation. These results highlight positive effects of DBS on cognitive functions that may benefit PD patients in daily-life association-learning situations.

Introduction

Making appropriate choices between distinct options in daily life (for example friend or foe, food or nonfood) is vital for optimal behavior and requires learning the causal relation between events, actions and their outcomes. Decisions about how best to respond in a situation are often guided by past learning, particularly when expectations about the outcomes of those decisions are well formed. In novel situations, expectations about the favorability of a decision's outcome (i.e., leads to reward vs. leads to punishment) are uncertain, and the associations between a situation, a response to it, and the outcome of that decision must be learned on the basis of trial and error.

Reward-based decision-learning paradigms enable us to measure the process of learning associations between stimuli, actions, and their related rewards. Several brain areas have been linked to key aspects of reward-based decision learning, including prefrontal regions (e.g., the dorsolateral and orbitofrontal cortices) and the basal ganglia. Among the latter structures, the subthalamic nucleus (STN) has been implicated recently as playing a key role in decision-making processes (Frank, 2007). The purpose of the present investigation was to determine the effects of STN stimulation on reward-based learning by testing the performance of patients with Parkinson's disease (PD) who had been treated with STN deep brain stimulation.

The Basal Ganglia in Reward-Based Decision-Learning

Although the basal ganglia are traditionally known to contribute to motor function, (Alexander, DeLong, & Strick, 1986; Alexander, Crutcher, & DeLong, 1990), contemporary views suggest involvement of the basal ganglia in several types of learning, including habit formation, procedural skill learning, and reward-based learning (Brown & Marsden, 1998; Kimura, 1995; Knowlton, Mangels, & Squire, 1996; Packard & Knowlton, 2002; Schultz, Tremblay, & Hollerman, 2003). Before discussing the role of the STN in such processes, we briefly turn to the prominent role of the striatum in reward-based decision-learning.

The Role of the Striatum in Reward-Based Decision-Learning

Lesion and human imaging studies demonstrate an important contribution of the striatum during reward-based decision-learning and support a functional dissociation between various aspects of the striatum (for an overview see Balleine, Delgado, & Hikosaka, 2007). The ventral striatum is more strongly associated with the establishment of expectations and motivations about the rewards of a response or decision. For instance, ventral striatal activity is commonly observed when actual rewards differ from expected rewards (i.e., reward-prediction error; Knutson et al., 2001; McClure et al., 2003; O'Doherty et al., 2004; Seger & Cincotta, 2005). In contrast, dorsal portions of the striatum are more strongly implicated in cognitive and motor aspects of reward-based learning (O'Doherty et al., 2004; Seger & Cincotta, 2005; Seymour, Daw, Dayan, Singer, & Dolan, 2007; Tricomi, Delgado & Fiez, 2004). For example, variations in dorsal striatal activity signal the evaluation of an action in terms of reinforcement and punishment. Furthermore, lesions to regions of the dorsal striatum as well as dopamine depletions in these areas disrupt formation of stimulus-response associations (Faure, Haberland, Condé, & El Massioui, 2005; Yin, Knowlton, & Balleine, 2004)

Although the dissociation between dorsal and ventral striatum is important to explain cognitive versus more affective aspects of learning, recent studies have suggested that a functional dissociation between two dorsal striatum structures, putamen and caudate, may account for these different aspects of learning. A dissociation between caudate and putamen contribution during reward-based learning was supported recently by a functional magnetic resonance imaging (fMRI) studies (Haruno & Kawato, 2006a, b). It was hypothesized that the caudate nucleus would be engaged in outcome evaluation processes during early phases of the learning task, whereas the formation of reward predictions based on stimulus-action-reward associations during later phases of the learning task would correspond more closely with activity in the putamen. These predictions were reasoned on the basis of differences in functional connectivity associated with the putamen and caudate nucleus; the putamen is embedded in the corticostriatal motor loop, whereas the caudate nucleus forms functional loops with the lateral orbitofrontal cortex (OFC) and the dorsolateral prefrontal cortex (DL-PFC).

Probabilistic learning tasks measure the evolvement of expectancies about the outcomes of a decision as subjects attempted to form stimulus-action-dependent reward predictions (SADRP; e.g., Haruno & Kawato, 2006a, b). For instance, in one such task (Haruno & Kawato, 2006a) subjects were instructed to maximize their total monetary reward by pressing a left or right button when a fractal stimulus appeared on the screen. In order to maximize their reward participants had to learn the probabilistic association between visual stimuli, responses and rewards. In the initial stages of learning, the associations are unknown. Thus, expectations about the potential reward of a decision are often disconfirmed by the actual reward. The difference between expected and actual rewards, coined the reward-prediction error (RPE), is theorized to provide the feedback necessary to adjust decision-making strategies. As learning progresses, stimulus-action-dependent rewards are predicted more proficiently: the SADRP is enhanced and the RPE is reduced as subjects more accurately anticipate the rewards associated with their actions.

Functional magnetic resonance imaging (fMRI) during performance of such probabilistic learning tasks showed that activation of the caudate nucleus (ventral and dorsal) and its associated circuitry (OFC and DL-PFC, involved in generating and testing hypotheses regarding reward optimization) closely corresponded with the RPE, especially in the early stages of learning (Haruno & Kawato, 2006a). The global reward-related features of these stimulus-action-reward associations are propagated from the caudate to motor loops (which include the putamen and premotor areas) by means of a dopamine signal (subserved by reciprocal projections between the striatum and the substantia nigra; Haruno & Kawato, 2006b). During later stages of learning, putamen activity increases with reward predictions (i.e., with learning SADRPs). Activity in the putamen increases to incorporate more specific motor information with the associated stimuli and the expected reward; that is, the reward associated with a specific stimulus and a specific action becomes more predictable and learning is gradually fine-tuned (Haruno & Kawato, 2006b). As these SADRP values increase, the RPE is reduced as subjects more accurately anticipate the rewards associated with their actions.

Note that the change in emphasis from RPE during early phases of learning to SADRP during later stages bears resemblance to the phasic DA bursts displayed by striatal neurons after unexpected reward during early phases which shift to the time of

conditioned reward-predicting stimuli during later stages (Balleine et al., 2007; Schultz et al., 2003).

Parkinson's Disease and Reward-Based Decision-Learning

Studies with patients PD have provided further support for the role of the basal ganglia in reward-based learning. PD is a neurodegenerative process commencing in the midbrain, in particular in those dopaminergic neurons of the substantia nigra that project in a compact bundle of fibers into the dorsolateral striatum (mostly the putamen; Bjorklundt & Dunnett, 2007). Thus, PD is initially characterized by dopamine depletions in the striatum that produce motor deficits, such as tremor, bradykinesia, and rigidity (McAuley, 2003); subsequently, these effects extend to cognitive deficits, such as impairments in reversal learning, decision-making, working memory, response inhibition, and speed/accuracy balancing (Cools et al., 2001, Cooper et al., 1992; Swainson et al., 2000; Wylie et al., 2009a, b, in press).

Dopaminergic projections within the striatum are differentially affected by Parkinson's disease. Early in the disease, dopamine is more severely depleted in the motor (including putamen and supplementary motor areas) and dorsolateral loops (including the DL-PFC and the dorsolateral head of the caudate) compared to the orbitofrontal (lateral OFC, ventromedial head of caudate) and anterior cingulate loops (anterior cingulate, ventral striatum; Kaasinen & Rinne, 2002).

The Role of the Subthalamic Nucleus in Reward-Based Decision Learning

Although the most common treatment for PD consists of DA precursors (typically levo-dopa) and agonists, deep brain stimulation (DBS) of the STN has become the treatment of choice in patients whose symptoms are less well controlled by medications (Lang, 2000; Limousin et al., 1995). While there is an ongoing debate about the specific mechanisms underlying therapeutic effects, the remedial effects of DBS on the motor symptoms of PD are substantial (Benabid, 2003; Benazzouz & Hallett, 2000; Bergman,

Wichmann, & DeLong, 1990; Liu, Postupna, Falkenberg, & Anderson, 2006; Meissner et al., 2005). PD patients treated with DBS of the STN for the purpose of alleviating motor symptoms afford the unique opportunity to investigate stimulation of a specific basal ganglia region on cognitive functions.

Substantial evidence from animal studies (Baunez et al., 2001; Bergman, Wichmann, Karmon, & DeLong, 1994; Karachi et al., 2004) and PD patient studies (Jahanshahi al., 2000; Schroeder et al., 2002; van den Wildenberg et al., 2006; Witt et al., 2004) documents that the STN is critically involved in both motor control and action selection (Boraud et al., 2002). The role of the STN and the effects of DBS of the STN on reward-based decision-learning processes has not been studied as extensively. The effects appear more variable, ranging from null effects to impairments in some studies and improvements in others (Frank, Samanta, Moustafa, & Sherman, 2007; Funkewiez et al., 2006; Jahanshahi et al., 2000; Saint-Cyr et al., 2000). The STN receives sensorimotor, cognitive and limbic input from the external segment of the globus pallidus (GPe). While these projections stem from functionally separate sources, the boundaries between sensorimotor, cognitive, and affective territories within the STN are not sharply defined (Karachi et al., 2005), nor is there a clear segregation between modalities in the output of the STN (Sato, Parent, Levesque, & Parent, 2000). Mallet et al. (2007) recently proposed that the STN not only regulates input from different modalities, but also integrates sensorimotor, emotional and cognitive aspects of behavior.

Additionally, the STN may receive reward-related information (i.e. expected magnitude of reward) from medial OFC projections to STN similar to the input from premotor cortex, as has been shown in rats (Maurice, Deniau, Glowinski, & Thierry , 1998) and hold the response output system (thalamus) in check until the expected reward options for a certain response are evaluated. Stimulating the STN may disinhibit the limbic circuits analogous to the disinhibition of motor circuits.

There is indeed some evidence from animal studies indicating that the STN plays a role in reward-based decision-making. STN lesions in rats have been shown to increase the incentive salience of reward-related stimuli (Uslaner & Robinson, 2006; Uslaner, Dell'Orco, Pevzner, & Robinson, 2008) which could be an indication of enhanced motivation and may affect learning by increasing reward motivation.

Studies on reversal learning, which depends mainly on learning from negative feedback, showed improved performance with STN stimulation in medication-withdrawn PD patients (Funkewiez et al., 2006) and in animals with STN lesion (El Massioui, Cheruel, Faure, & Conde., 2007). In contrast, learning based on positive and negative feedback in a reward-based probabilistic learning task remained unchanged in mildly medicated PD patients ON compared to OFF STN stimulation (Frank, 2007). Associative learning (stimulus-response learning not based on reward or action outcome) was in fact observed to decline in PD patients treated by DBS of the STN, either with or without levo-dopa medication (Jahanshahi al., 2000; Saint-Cyr, Trépanier, Kumar, Lozano, & Lang, 2000).

Since previous studies with STN stimulation in PD patients and STN lesions in animals produced such a mixed set of results, the present experiment sought to further examine the influence of STN stimulation on RPE processing and the formation of SADRP among PD patients.

Present study

The current study investigated the effect of STN stimulation on reward-based decision-learning. Patients with PD performed the Haruno and Kawato (2006a) task with their DBS device turned OFF and with the device turned ON, with the order of DBS counterbalanced across subjects. We determined the effect of DBS of the STN on RPEs during the early phase of learning and on formation of SADRPs during the last phase of learning. DBS is targeted primarily at the motor regions of the STN, and such stimulation clearly enhances motor functions (Kleiner-Fisman et al., 2003), which are supported primarily by regions of the dorsal putamen and its associated circuitry in the motor loop. Therefore, we predicted that STN stimulation has beneficial effects on the formation of putamen-based SADRP values during later stages of learning. Stimulating the motor area of the STN may affect cognitive and limbic loops as well, because STN output is not sharply segregated (Mallet et al., 2007). Hence, DBS of the STN might also serve to reduce caudate-based RPE values during earlier stages of learning, although such an effect would be less direct and perhaps less potent.

Methods

Table 1

PD patient information.

Variable	**Sample N = 12**	
	Mean	**SE**
Age (yrs)	61.1	2.3
Sex (male/female)	9/3	
MMSE	29.1	0.3
Finger tapping ON (# taps)	42	2
Finger tapping OFF (# taps)	34	3
Pegboard ON (time in sec)	31.4	2.5
Pegboard OFF (time in sec)	34.6	3.2
Years since disease onset	11.4	1.8
L-Dopa (daily dose mg)	425.0	81.5
Stimulation details		
Left		
Voltage (V)	3.2	0.2
Rate (Hz)	138.2	4.2
Pulse Width (microsec)	68.2	4.2
Right		
Voltage (V)	3.1	0.3
Rate (Hz)	138.1	4.7
Pulse Width (microsec)	73.6	4.2

Patients and surgery details

Our study included 12 PD patients who were treated successfully with bilateral DBS of the STN and who were concurrently treated with dopaminergic medications (with the

exception of one patient). Patients remained on their prescribed doses of dopamine medications during their testing and were studied during their optimal therapeutic window. Table 1 shows participant information.

All participants were free of dementia and did not demonstrate clinical levels of depression at the time of testing. Participants with PD were recruited from the Movement Disorders and DBS Neurosurgery clinics at the University of Virginia Medical Center and diagnosed with PD by a neurologist specializing in movement disorders. All PD patients ambulated independently and were rated a Hoehn & Yahr Stage III or less when their DBS device was turned on.

The surgical placement of DBS macroelectrodes followed standard procedures described previously (Elias, Fu, & Frysinger, 2007; De Salles et al., 2004). Briefly, stereotactic surgical procedures were performed with the Leksell model G stereotactic frame and arc. Frame placement occurred the morning of surgery, and preoperative MRI was obtained immediately following frame placement. Procedures were performed under local anesthesia without sedation to permit intraoperative stimulation testing. Electrodes were inserted through 14-mm-diameter precoronal bur holes. A Navigus intracranial cap (Image-Guided Neurologics) was used to secure the DBS electrodes. Macroelectrodes (Medtronic Model 3389) consisting of 4 platinum-iridium cylindrical surfaces were used, each with a diameter of 1.27 mm, length of 1.5 mm, and edge-to-edge separation of 0.5 mm. Macroelectrodes were guided into the motor area of STN using MRI-guided stereotaxy and intraoperative microelectrode recordings. The planned coordinates for macroelectrode placement was based on direct visualization of the STN on T2-weighted MR images. Final electrode position was based on microelectrode recordings and confirmed intraoperatively with macrostimulation after implantation of the DBS electrode. Anatomic position of the electrode was determined postoperatively with MRI 3D reconstructions.

The PD patients completed participation under two conditions, once with their DBS device turned off and once with the device turned on. Patients were tested at a minimum of 3 months post-surgery. Exclusion criteria included the following: dementia; history of neurological condition other than PD; untreated or unstable mood disorder; history of bipolar affective disorder, schizophrenia, or other psychiatric condition known to compromise executive cognitive functioning; and untreated or unstable medical condition known to interfere with cognitive functioning (e.g.,

diabetes, pulmonary disease). All participants had normal or corrected-to-normal vision. All subjects participated voluntarily and gave their written informed consent prior to participation, as part of procedures that complied fully with relevant laws and with standards of ethical conduct in human research as regulated by the University of Virginia human investigation committee.

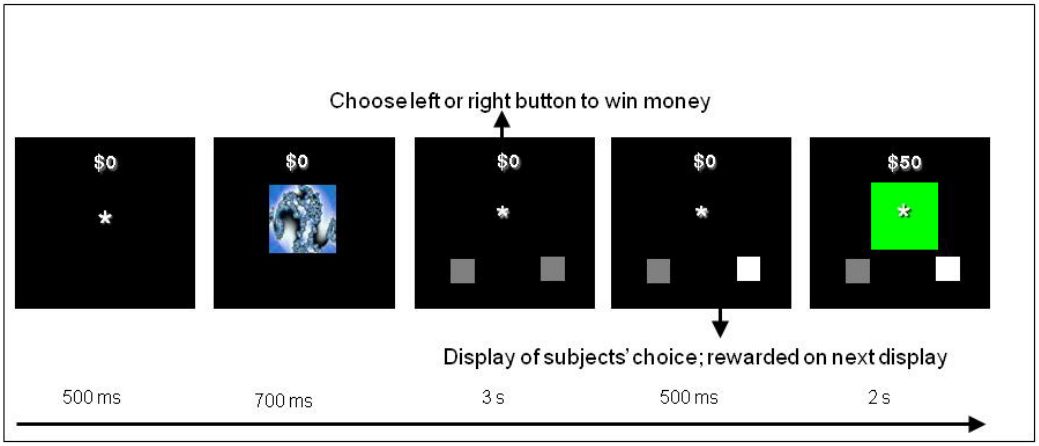
Task and apparatus

Questionnaires

The mini-mental status examination (MMSE, Folstein, Folstein, & McHugh, 1975) assessed the global cognitive state of each patient to verify the absence of dementia (i.e. MMSE score higher than 25). To capture the effects of DBS of the STN on fine motor dexterity and speed, we administered the Purdue Pegboard task (Lezak, 1995) and a finger-tapping test during each session. The latter required participants to use the index finger of each hand to tap a tapping board as fast as possible during a period of 10 seconds. The tapping task alternated between each hand until three attempts were completed with each hand.

Figure 1

Trial example of the probabilistic learning task adapted from Haruno and Kawato (2006a). In the example, the subject receives a reward by pressing the right button with this specific stimulus.



Experimental paradigm

A probabilistic learning task, adapted from Haruno and Kawato (2006a, 2006b), was implemented on an IBM-compatible computer with a 17-inch digital display monitor. The computer screen, placed at a distance of 91 cm, was positioned so that stimuli appeared at eye level. Stimuli consisted of colored pictures against a dark background. Responses to stimuli were right or left thumb button presses registered by comfortable handheld grips.

The probabilistic learning task was designed to estimate RPEs and measure the learning of SADRPs which have been linked to caudate nucleus and putamen activity, respectively. Subjects were instructed that the goal of the task was to make as much money as possible by pressing a left or a right button press to each picture stimulus that appeared on the computer screen. Each response provided the chance to either win or lose $50. This was introduced as game-money; participants were not remunerated for their participation. Figure 1 depicts the sequence of a trial from the task. Each trial began with the presentation of a fixation point (an asterisk) in the center of the screen, and subjects were instructed to focus on this point in anticipation of the presentation of

a picture stimulus. After a duration of 500 ms, the fixation point was extinguished and one of three picture stimuli (colored fractals) appeared in the same location as the fixation point. The picture stimulus subtended visual angles of 5.67° x 4.41° (9 x 7 cm). The picture stimulus remained on the screen for 700 ms. Participants were instructed to view the picture stimulus, but not to respond until the picture stimulus disappeared and was replaced by a response screen. The response screen consisted of the fixation point and two gray boxes displayed at the bottom left and bottom right portions of the screen, respectively (see Figure 1). Upon the presentation of the response screen, the participant was instructed to make a left or a right button press, which would then be indicated on the screen by a change in color (from gray to yellow) of the box that corresponded to the response side that was chosen (left button press = left box turns yellow). The participant was given 3 seconds to issue a response. After the button press was indicated on the screen, a large box with feedback appeared in the center of the screen. If the participant chose the correct response, the large box appeared in green, indicating that $50 had been won. If the incorrect response was chosen, the box appeared in red, indicating that the participant had lost $50. Throughout the entire trial, a running tab of the total amount of money won by the participant was depicted in the upper center portion of the screen. Thus, if the participant won or lost $50 on a particular trial, the running total was immediately updated.

Subjects completed three sessions of 48 trials. For each session, a novel set of 3 picture stimuli were used. The reward outcome of each response to a picture stimulus was determined in the following way: (1) for each picture, one response hand was assigned as the optimal choice and the other response hand was designated as the nonoptimal choice; (2) in the first session selecting the optimal response hand resulted in a 90% probability of winning $50 and a 10% probability of losing $50; (3) in a second session, selecting the optimal response hand resulted in an 80% probability of winning $50 and a 20% probability of losing $50; (4) in a third session selecting the optimal response hand resulted in an 70% probability of winning $50 and a 30% probability of losing $50. In all sessions, the probabilities of winning versus losing were reversed for the nonoptimal relative to the optimal response hand. As an example, in the 90/10 session a left response to fractal stimulus 1 (FS1) yielded a 50 dollar reward with a probability of 0.9 (90%) and a 50 dollar loss with a probability of 0.1 (10%). A right response to FS1 yielded a 50 dollar loss with a probability of 0.9 and a 50 reward with a

probability of 0.1. Therefore, the optimal behavior for FS1 in the 90/10 session was to press the left button, which participants had to learn by trial and error. The dominant probabilities for optimal behavior regarding the other fractal stimuli (FS2 and FS3) in the 90/10 session were also 0.9. The optimal response for each fractal was pseudo randomized over left and right hands, for example optimal behavior could be FS1: right, FS2: left, FS3: right, which means that these responses were rewarded with positive feedback 90% of the time. Similarly, a response pattern could consist of two fractals that were rewarded (most of the time) with a left hand response and one with a right hand response. For each session, one of these two response patterns was randomly and the specific response options were randomly attached to each of the fractals. Additionally, the fractal stimuli were randomly presented and occurred equally frequent within a session. Session order was randomized as well.

Procedure

Participants completed two similar versions of the task on the same day. The versions were similar in all respects except the picture stimuli differed. The task was completed with the patient's DBS device turned on (ON condition), and again with the DBS device turned off (OFF condition). The order of testing with respect to the status of the DBS device was counterbalanced and randomly determined among patients. Prior to completing the task, each participant signed the consent form and completed the MMSE. As well, each participant completed the pegboard and finger tapping tasks with the DBS device turned on and turned off. After turning the DBS device ON or OFF, patients waited 30 minutes before commencing the cognitive task.

Data Analysis

Computational model to estimate SADRP & RPE

A reinforcement model (Q-learning, Sutton & Barto, 1998) was used to estimate each participant's SADRP and RPE during learning. Q-learning is an implementation of a temporal difference model which assumes that stimulus-action-reward associations are

acquired as a single representation during learning. The SADRP value (Q) consists of the predicted amount of reward for a certain decision (left or right response, r) made for a specific stimulus (one of three fractal stimuli, FS). It thus relates reward to sensory input and actions. Individual predicted reward values (SADRPs) for each action (two response) and each fractal stimuli (three different fractal stimuli) will be calculated at time t, Q_t (FS, r) which adds up to six SADRP values per session.

The RPE represents the actual reward received (rt) minus the expected reward, RPE = rt-(Q_t (FS, r). For the next occurrence of the same stimulus and action, SADRP and RPE values are updated according to the "Q-learning algorithm" to maximize reward (Sutton and Barto, 1998), Q_{t+1}(FS, r) = Q_t (FS, r) + a_t^{FS} (rt-(Q_t (FS, r))).

The learning rate is updated separately for each FS according to the following rule: $a_t^{FS} = (a_{t-1}^{FS})/ (1+ a_{t-1}^{FS})$.

The formula of this learning rate is often used in reinforcement learning studies or studies on adaptive control (Bertsekas & Tsitsiklis, 1996; Dayan, Kakade, & Montague, 2000; Haruno & Kawato, 2006a, 2006b; Young, 1984). It provides an estimation of a learning parameter which is updated recurrently with the presentation of a stimulus. In the current study, a_t^{FS} reduces with the presentation of each fractal stimulus, but remains equal if a specific FS is not presented.

The learning rate (a_t^{FS}) decreases towards the end of the learning stage (when SADRP becomes reliable). This is an important feature of a_t^{FS} because it means that, at the end of learning, the SADRP is less affected by an unexpected RPE (due to the probabilistic nature of the task).

The RPE is large at the beginning of learning (i.e. first 24 trials), while the SADRP value is small. Major changes in SADRP are especially expected in the first stage of learning. In a later stage of learning (i.e. last 24 trials) SADRP becomes accurate and does not show large changes (converges to an asymptotic value). Additionally, RPEs are expected to be small at the end of learning.

Analyses

Motor performance on the finger tapping and pegboard tests was analyzed separately by a one-tailed paired samples t-test. We expected motor performance to improve with the DBS device turned ON compared to when it was turned OFF. A one-sample t-test

was used to test whether MMSE scores (OFF stimulation) were significantly larger than 25.

SADRP and RPE values were analyzed separately by Repeated-Measures analysis of variance (ANOVA), including the within-subjects variables of Stimulation (OFF, ON) and Session (90/10, 80/20, 70/30). Session types are represented as the dominant versus nondominant probability. Specific predictions were tested by using a Simple Contrast test; that is, the 90/10 session was compared with the 80/20 session and the 70/30 session. The analyses were based on the mean RPE value (calculated on the first 24 trials) and the mean SADRP value (based on the last 24 trials).

Since individual disease characteristics of PD patients may affect cognitive performance, like disease duration and age (Kaasinen & Rinne, 2002), we also took these variables into account. Disease duration and age were correlated with the dependent variables (improvement in RPE and SADRP comparing ON and OFF DBS) to identify which individual characteristics could be predictive for performance in the learning task. The variables that turned out to correlate significantly with the dependent variables were used as predictors in the subsequent regression analysis.

First, we correlated change in RPE (RPE ON minus OFF = ΔRPE) and change in SADRP (SADRP ON minus OFF = ΔSADRP), separately for each session, with individual characteristics (disease duration and age). Note that small RPE values are expected ON stimulation and high RPE values OFF stimulation. Thus negative ΔRPE indicates that participants improved, whereas positive ΔRPE indicates that they were impaired ON compared to OFF stimulation. SADRP values are expected to increase ON versus OFF stimulation; therefore high ΔSADRP indicates improved performance.

A stepwise regression analysis was then performed with the variables that turned out to significantly correlate with RPE and SADRP, that is, disease duration and age with Δ RPE in the 90/10 session. ΔSADRP did not significantly correlate with any of the individual characteristics. Thus, dependent variables in the regression analysis were change in RPE ON compared to OFF (ΔRPE in session 90/10) and independent variables consisted of disease duration and age.

Results

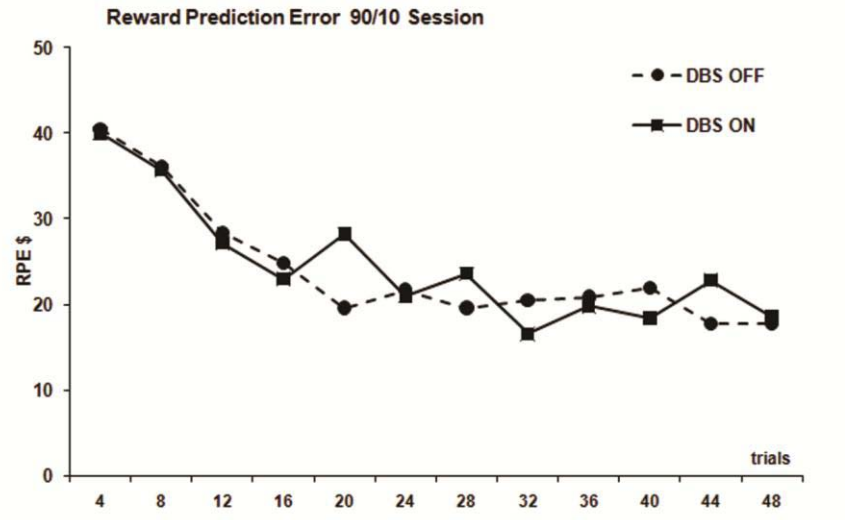
Motor performance

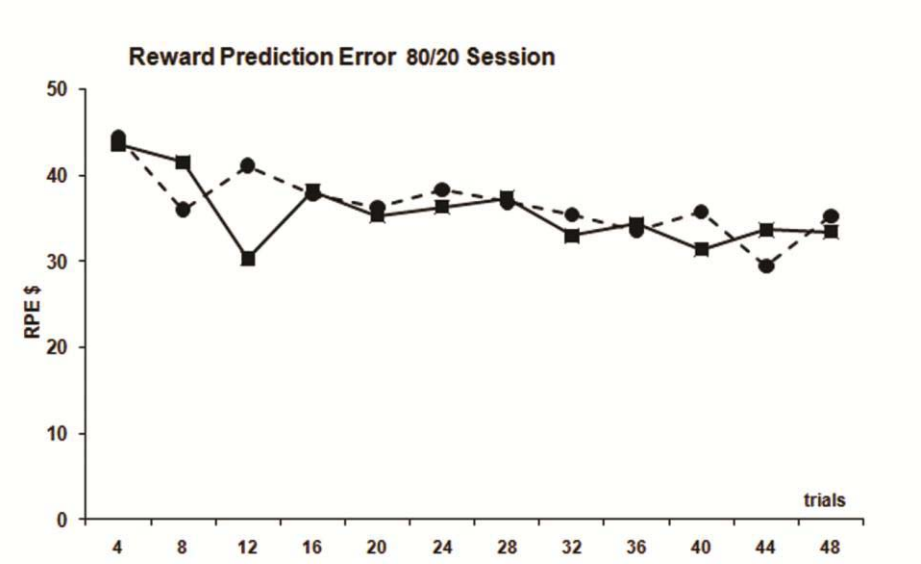
Consistent with improved fine motor control associated with DBS of the STN, turning the DBS device ON (compared to turning the device OFF) increased finger tapping speed, t (11) = 3.5, $p < 0.01$, and nearly significantly sped pegboard performance t (11) = -1.7, p = 0.06. MMSE scores OFF stimulation were significantly larger than 25, M = 29.1, t (11) = 12.2, $p < 0.001$, indicating that our participants were not demented.

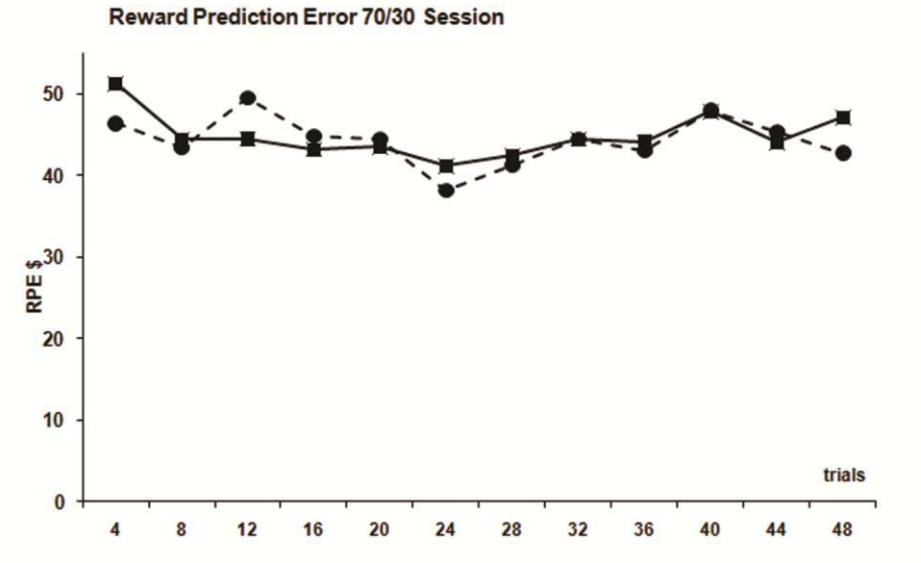
Figure 2 shows the course of mean RPE and SADRP values ON and OFF stimulation during the task (for each of the sessions, i.e. 90/10, 80/20, 70/30). Figure 3 displays the mean RPE values from the first part of each session and the mean SADRP values from the second part of each session.

Figure 2.

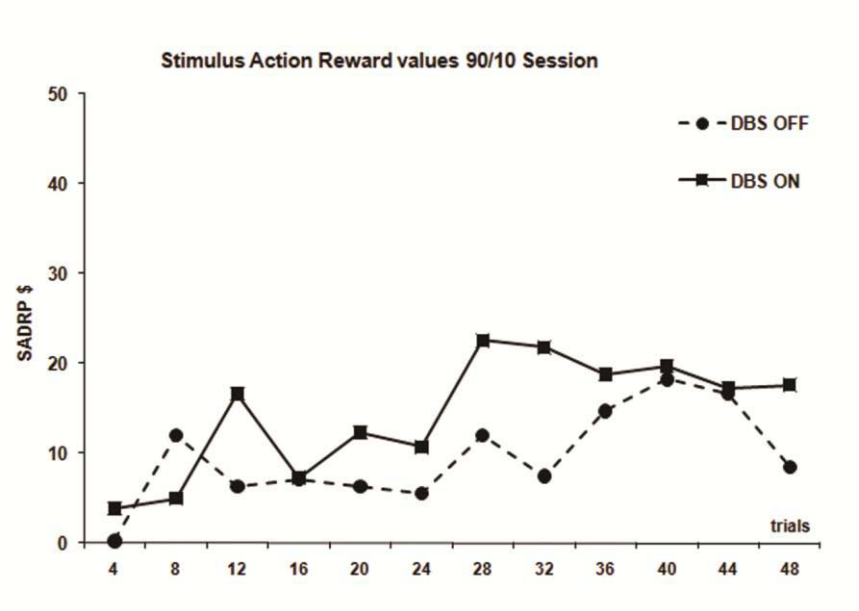
a. Trial-by-trial course of mean RPE values ON and OFF stimulation, separate for each session.

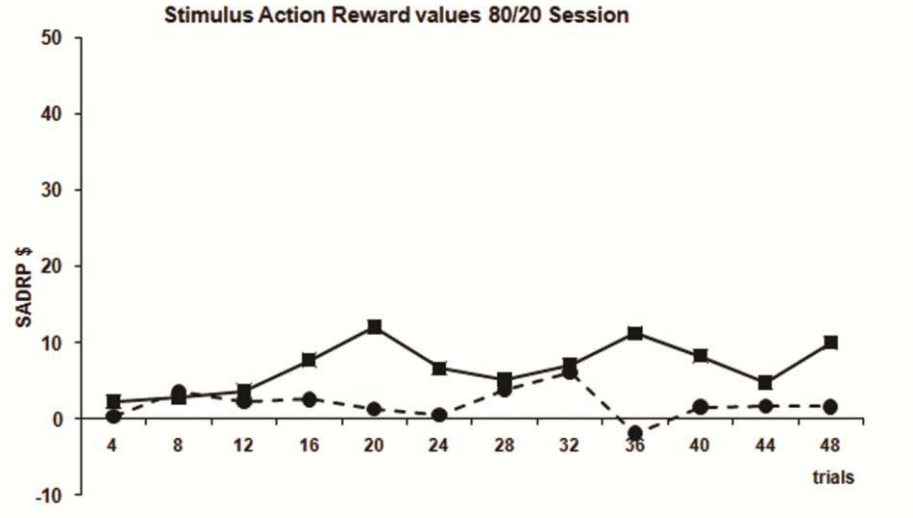



Reward Prediction Error 80/20 Session
RPE $
trials
0
10
20
30
40
50
4
8
12
16
20
24
28
32
36
40
44
48


Reward Prediction Error 70/30 Session
RPE $
trials
0
10
20
30
40
50
4
8
12
16
20
24
28
32
36
40
44
48

b. Trial-by-trial course of mean SADRP values ON and OFF stimulation, separate for each session.





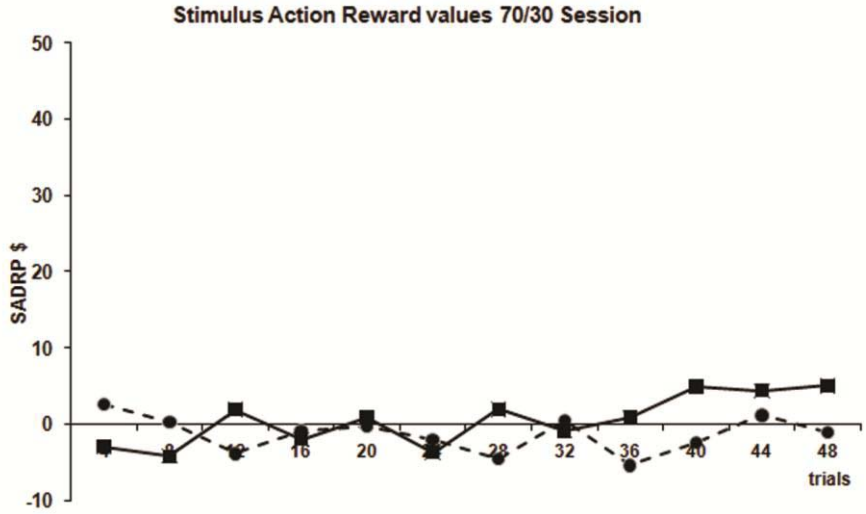
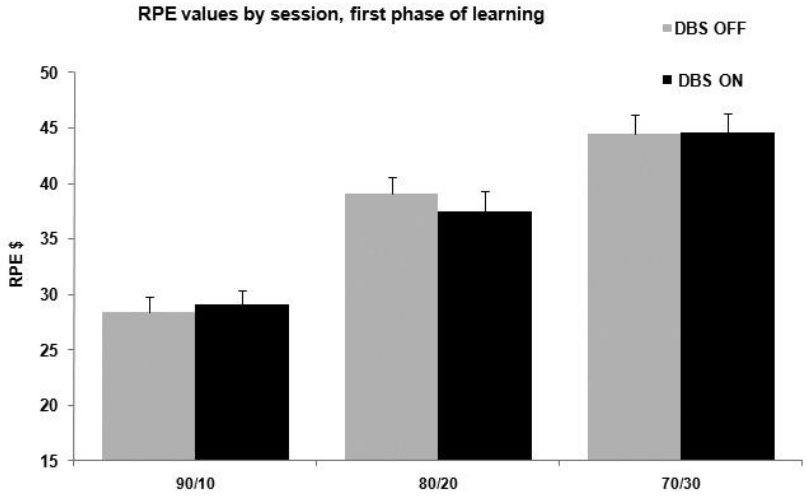
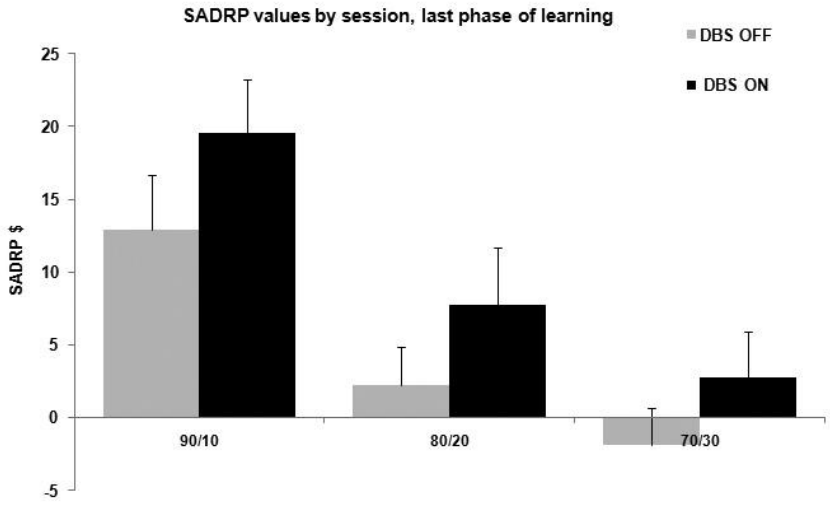

Stimulus Action Reward values 70/30 Session
SADRP $
50
40
30
20
10
0
-10
16
20
32
36
40
44
48
trials

Figure 3

a. Mean RPE values from the first 24 trials separate for each session.



b. Mean SADRP values from the second 24 trials separate for each session.



Reward-Based Decision-Learning

RPE

RPE values varied across Sessions, F (2, 22) = 42.17, p < 0.001. Planned contrasts revealed a smaller RPE in the 90/10 Session ($M_{90/10}$ = 28.76) compared to the 80/20 and 70/30 Session, $F_{90/10\text{-}80/20}$ (1, 11) = 25.96, p < 0.001, $M_{80/20}$ = 38.26; $F_{90/10\text{-}70/30}$ (1, 11) = 79.60, p < 0.001, $M_{70/30}$ = 44.54 (Figures 2a and 3a). DBS of the STN did not influence RPE, F (1, 11) = 0.03. Moreover, RPE did not vary as a function of the interaction between Session and DBS, F (2, 22) = 0.46. These findings suggest that DBS of the STN did not influence reward-prediction error processing related to the caudate early in learning.

SADRP

The learning of associations depended on the probability of being rewarded for a correct response (i.e., the Session effect), F (2, 22) = 17.36, p < 0.001. SADRP values in the 90/10 Session ($M_{90/10}$ = 16.23) were significantly larger than values obtained in the 80/20 and 70/30 Session, $F_{90/10\text{-}80/20}$ (1, 11) = 23.64, p < 0.01, $M_{80/20}$ = 5.00; $F_{90/10\text{-}70/30}$ (1, 11) = 32.76, p < 0.001, $M_{70/30}$ = 0.40, see Figure 3b. Thus, patients learned better when the correct action was more likely to be rewarded. In contrast to the analysis of RPE, DBS of the STN benefitted the learning of SADRPs, F (1, 11) = 8.11, p < 0.05. Specifically, participants ON DBS showed significantly larger SADRP values (M_{ON} = 10.00) than patients OFF DBS (M_{OFF} = 4.414), see Figure 2b. No significant interaction between Stimulation and Session was found, F (2, 22) = 0.05, suggesting that DBS of the STN improved performance in all sessions equally.

Table 2

ΔSADRP and ΔRPE ON compared to OFF STN stimulation correlated with disease duration and age.

Variables	**Disease duration**	**Age**
Disease duration(yrs)	1.00	0.15
Age (yrs)	0.15	1.00
ΔRPE 90/10 session	0.76**	0.51
ΔRPE 80/20 session	-0.16	0.05
ΔRPE 70/30 session	0.03	0.02
ΔSADRP 90/10 session	-0.40	-0.06
ΔSADRP 80/20 session	0.28	0.27
ΔSADRP 70/30 session	-0.20	0.08

**p<0.05 **p<0.01*

Table 3

Linear stepwise regression on ΔRPE ON-OFF in the 90/10 session as a function of disease duration and age.

Variables	**β**	**R**	**ΔR^2**
Step 1 Disease duration	0.70	0.76	0.58
Step 2 Age	0.41	0.87	0.16

Relation ΔSADRP and ΔRPE ON-OFF with disease duration and age

Although the ANOVA did not show an effect of DBS of the STN on RPE values, correlation analyses revealed that the change in RPE between DBS conditions was sensitive to individual characteristics. ΔRPE ON-OFF DBS within the 90/10 Session correlated significantly with disease duration, $r = 0.76$, $p < 0.01$ and showed a large correlation with age, $r = 0.51$, although this was not significant, $p = 0.09$. Thus, patients who were younger and earlier in the course of the disease showed the largest improvement in the RPEs during the initial stages of learning when their DBS device was turned ON.

In contrast, ΔSADRP ON-OFF DBS within the 90/10 Session did not correlate with age or disease duration (see Table 2 for correlations). RPE and SADRP values of the 80/20 and 70/30 Session did not reveal any significant correlations with disease duration, or age (see Table 2 for correlations) and were thus not included in the regression analysis.

Stepwise regression, with disease duration and age entered sequentially as predictors, showed disease duration to be a significant predictor of ΔRPE, $F(1,11) = 14.06$, $p < 0.01$, with age also explaining additional variance, $F(2, 11) = 13.39$, $p < 0.01$. These effects are presented in Table 3 for R and ΔR^2 for each of the predictive variables.

Discussion

The present study sought to investigate the effect of STN stimulation on separate components of reward-based learning: outcome evaluation (the processing of reward-prediction errors to update hypotheses) and reward anticipation (the formation of SADRPs) that have been tied to distinct regions in the striatum and their associated circuitries. The probabilistic reward-based decision-learning task used here successfully reproduced the behavioral findings first reported by Haruno and Kawato (2006a; 2006b). Specifically, participants were able to adapt to RPEs made during early stages of learning and showed increased learning of SADRPs across trials, especially in the

condition with the highest degree of reward-related predictability. Similar to Haruno and Kawato (2006a), the conditions with a lower degree of reward-predictability turned out to be more difficult; performance dropped dramatically in patients ON as well as OFF DBS.

PD patients completed the task twice, once with their DBS device turned ON and again with the device turned OFF. We predicted that STN stimulation would improve the formation of stimulus-action-reward associations as reflected in higher SADRP values. Consistent with this prediction, SADRP at the late stages of learning was larger when STN stimulation was turned ON compared to when it was turned OFF. Because SADRP values have been linked to activity in the putamen, this finding provides indirect support for the idea that DBS of the STN may benefit the action-oriented learning functions of the severely dopamine-depleted putamen in PD patients. This finding fits well with studies of PD showing that DBS of the STN enhances motor performance and control (Benabid et al., 1993; Kleiner-Fisman et al., 2003; van den Wildenberg et al., 2006) and improves reward-based learning (Funkiewiez et al., 2006) It is also consistent with studies of STN lesions in rats (note that stimulation of the hyperactive STN in human PD patients is assumed to lead to roughly comparable inactivation effects as lesioning the STN in animals), indicating that STN lesions increase 'wanting' and thereby facilitate reward-based learning (Uslaner & Robinson, 2006; Uslaner et al., 2008), particularly when the probability of receiving positive reward is high. The modulating effect of STN was indeed most salient in the highly predictive session (dominant probability 0.9).

In addition to determining the effects of DBS of the STN on learning proficiency, we also considered its effects on the processing of reward-prediction errors that occur in the early stages of learning. For learning to be successful, subjects must evaluate discrepancies between expected (or predicted) rewards associated with a particular decision and the actual outcome of that decision. When an error occurs (i.e., predicted reward does not match the actual outcome), expectancies about possible outcomes associated with a decision can be updated to increase the likelihood of selecting a more optimal (i.e., reward-yielding) response in the future. As expectancies about the outcomes of particular decisions become more accurate, subjects are less swayed by the occasional violation of these reward expectancies and learn to optimize their selection of the most advantageous response to a stimulus. Thus, the processing of RPEs,

especially early in the course of the learning experience, is a fundamental aspect of effective learning. Overall, DBS of the STN did not influence RPE values. This suggests that the processing of reward-prediction errors, which has been linked to caudate nucleus activity (part of the cognitive corticostriatal loop), is insensitive to STN modulation, which contrasts with our prediction that stimulating the motor area of the STN would also affect cognitive and limbic loops (Mallet et al., 2007).

However, an interesting finding concerning RPE values emerged when we took into consideration individual differences within the PD group. Specifically, younger patients with relatively short disease duration did show improvement in RPE values when their DBS was turned ON compared to when it was turned OFF. The reason for this association is unclear, although interestingly another study reported beneficial effects of DBS of the STN on aspects of learning among PD patients who were younger (mean age 54.5 years, SD = 7.5) and who had a shorter disease duration (mean disease duration 10.7 years, SD = 3.9) (Funkiewiez, 2006). Similarly, several clinical studies reported that younger patients and patients with a relatively short disease duration benefit more from stimulation of the STN in terms of general motor performance than older patients and patients who had a longer disease duration (Charles et al., 2002; Pahwa et al., 2006; Schupbach et al., 2007).

Using a different probabilistic learning task, Frank and colleagues (Frank et al., 2007) failed to observe effects of STN stimulation on either positive or negative feedback learning in PD patients. According to a neurocomputational model developed to simulate behavior on their task (Frank, 2005), the STN provides a global NoGo signal because projections from the STN to GPi are diffuse and not response specific. Thus STN stimulation was predicted to exert no effects on learning specific stimulus-response associations. However, Frank et al. (2007) studied a group of PD patients who were on average older (mean age = 63.5, SD = 3.0) and who had a longer disease duration (mean = 14.8, SD = 1.65). The association between learning parameters and either age or disease duration was not investigated. Moreover, in our study, STN stimulation especially enhanced performance in the most predictive session; like in our less predictive sessions, in Frank et al.'s probabilistic learning task, the dominant probabilities (0.8, 0.7, 0.6) may have been less likely to show learning improvements by DBS of the STN.

Yet, the exact mechanism that established the STN stimulation effect remains unclear. In contrast with Frank's (2005) model, other BG models (e.g. Albin, Young, & Penney, 1989) would have predicted that stimulating the STN in PD patients would impair NoGo learning but improve Go learning. Stimulating the STN might have reduced the excessive activity in the NoGo pathway in PD patients in our study and thereby improved SADRP learning.

DBS of the STN in our study clearly improved the learning function associated with the putamen (SADRP) which might be explained by the placement of the stimulating electrodes in the motor areas of STN. Although STN output is not sharply segregated (Mallet et al., 2007) and stimulating the motor area of the STN may thus also affect cognitive and limbic loops, this DBS effect may nevertheless be relatively stronger in regions associated with the motor loop.

The beneficial effect of STN stimulation on the putamen may have been established by STN influence on multiple sites within the motor loop. STN stimulation may have modulated the processing of motor input information from GPe (entering the GPe via the putamen). Moreover, STN is directly activated by projections from the motor cortex (hyperdirect pathway, Nambu et al, 2000). Thus, if several competing responses are active in the motor cortex, the STN becomes increasingly activated which leads to a global NoGo signal. Stimulating the STN may change the way these signals are processed, for example, if an already overactive STN in PD is excited by the motor cortex this leads to oscillatory activity and tremor, whereas stimulating or lesioning the STN normalizes this activity (Bergman et al., 1990). Parametric modulation of STN stimulation in different functional STN areas might shed further light on the modulating role of STN in reward-based learning.

To summarize, our data suggest that the STN plays a modulatory role in reward based learning. STN stimulation modulated S-R learning and was associated with more efficient reward processing when clinical characteristics were taken into account.

Relation to other studies

In the current study feedback-based response selection was improved by STN stimulation. This is in line with the finding that action selection improves with STN

stimulation (van den Wildenberg et al., 2006), but in contrast with findings from rat studies indicating that STN lesions induce impulsive responding (Baunez et al., 2001; Baunez & Robbins, 1997). In the probabilistic learning task used in the current study, impulsive behavior would have led to less effective feedback processing and more random choices, which we did not find.

Other studies revealed that STN stimulation in PD patient reduced performance especially with conflict trials and when a speeded response is required such as in a Stroop task (Schroeder et al., 2002; Witt et al., 2004; Jahanshahi al., 2000). Note that our task did not emphasize speeded responding nor did it consist of stimuli inducing response competition (conflict trials). The results from these Stroop tasks, however, do not unanimously support the notion that STN stimulation impairs action selection with speeded conflict trials because some find prolonged reaction times but no effect on accuracy (Schroeder et al., 2002), while others only report more errors with Stroop conflict trials (Witt et al., 2004; Jahanshahi al., 2000).

Limitations

There are some limitations related to the experimental paradigm and thereby the interpretation of the results. Although SADRP and RPE have been linked to the role of DA bursts at different time points and in different stages of learning, and have been shown to correlate with different striatal structures, at the behavioral level they are not entirely independent. That is, a decrease in RPE values increases SADRP values (according to the computational model). Thus, a null result of stimulation status on RPE values at the beginning of the task but an effect on SADRP at the end of learning does not entirely exclude that the caudate is modulated by STN stimulation. Rather, it suggests that the STN stimulation does not affect learning in an early stage.

Currently it is unknown how many trials (and feedback) are needed to activate the caudate and putamen in PD patients and in what way this is modulated by STN stimulation, although there is some evidence that PD patients need more trials to learn (Shohamy, Myers, Kalanithi, & Gluck, 2008). Future studies should test the critical time

course of caudate and putamen involvement in probabilistic reward-based decision-learning in PD by means of an fMRI study.

Finally, the PD patients in our study remained on their regular DA medication, although these dosages are smaller than in regular medicated PD patients. Nevertheless, DBS of the STN affects reward-based decision-learning above and beyond a DA effect. Future studies that consider the medication and DBS effects separately as well as their interaction will be important.

Conclusion

In conclusion, DBS of the STN for treatment of PD motor symptoms also has a beneficial effect on learning stimulus-action-reward associations, a process shown previously to be associated with putamen activity. Thus, with DBS of the STN, PD patients were more effective at using feedback from their decisions to guide learning about how to respond optimally to a stimulus situation. Moreover, relatively young patients with shorter disease duration were particularly improved by DBS of the STN in their processing of reward errors early in the course of learning, which is essential for guiding new learning.