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

Wouwe, N.C. van

Chapter 5

Learning processes associated with caudate and putamen are improved by dopaminergic modulation Evidence from Parkinson's disease

Abstract

Learning to select optimal behavior in new and uncertain situations is a crucial aspect of living and requires the ability to quickly associate stimuli and actions that receive reward. Several studies converge on the notion that the striatum and dopamine (DA) are involved in this learning process. However, the modulatory role of DA in substructures of the striatum during reward-based learning, like the caudate and putamen, is not well established yet. The current study aimed to differentiate the influence of DA medication on reward processing and the formation of stimulus-action associations in patients with Parkinson's disease (PD) Patients with PD performed a probabilistic reward-based learning task while ON DA medication and while OFF DA medication. We determined the effect of medication on reward-prediction errors (RPE) during the early phase of learning and on the formation of stimulus-action-dependent reward predictions (SADRP) during the last phase of learning. RPE and SADRP have been tied to caudate and putamen activity, respectively, in a recent fMRI study (Haruno & Kawato, 2006a). Both reward-related processing and SADRP learning were influenced by DA medication. This findings suggests that the contributions of caudate and putamen activity to distinguishable aspects of reward-based learning are modulated by

DA activity. Moreover, DA pharmacotherapy for treatment of PD improves key elements of reward-based learning.

Introduction

Learning to adapt behavior and decision-making in a given circumstance to maximize reward is a fundamental aspect of living. For example, in new and uncertain situations, the ability to quickly acquire associations between stimuli and actions that receive reward ameliorates selection of optimal behavior. Reward-based decision-learning paradigms enable us to measure the process of learning associations between stimuli, actions, and their related rewards. Several brain circuits are involved in reward-based decision-making and -learning, including prefrontal cortices and subcortical areas like the basal ganglia (BG). Additionally, the neurotransmitter dopamine (DA) plays a modulatory role in these functions through projections from midbrain DA nuclei to the BG and cortical areas (Schultz, 2002). The current study aims to differentiate the role of DA in substructures of the striatum during reward-based decision-learning by means of testing patients with PD ON and OFF their DA medication on a probabilistic learning task.

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

Although the BG are traditionally known to contribute to motor function (Alexander, DeLong & Strick, 1986; Alexander, Crutcher & DeLong, 1990), more recently the BG have been shown to be engaged in several types of learning, including habit formation, procedural skill learning, and to reward-based decision-learning (Brown & Marsden, 1998; Kimura, 1995; Knowlton, Mangels, & Squire, 1996; Packard & Knowlton, 2002; Schultz, Tremblay, & Hollerman, 2003).

The BG consist of the striatum, the globus pallidus, the substantia nigra and the subthalamic nucleus. The striatum is the main input structure and receives projections

from the cortex and subcortial areas, i.e. DA projections from the substantia nigra and the ventral tegmental area (Mink, 1996). The striatum is connected with cortical brain areas in multiple cortico-striatal loops (Alexander et al., 1986, 1990); the motor, oculomotor, orbitofrontal, dorsolateral and anterior cingulate loops, subserving motor control, cognitive and motivational behavior. Because each of these loops serves a different function, the striatum has been posed to integrate motivational, cognitive and motor control information.

Lesion and human imaging studies demonstrate an important contribution of the striatum to 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 establishing expectations and motivational incentives with respect to 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, Fong, & Hommer, 2001; McClure, Berns, & Montague, 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, & Doyan, 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 DA depletions in these areas disrupt formation of stimulus-response associations (Faure, Haberland, Condé, & El Massioui, 2005; Yin, Knowlton, & Balleine, 2004).

Neuroanatomically, the dorsal striatum can be divided into the caudate nucleus and the putamen. The putamen and caudate nucleus are located in different corticostriatal loops; the putamen is embedded in the motor loop while the caudate nucleus is connected to the lateral orbitofrontal cortex and the dorsolateral prefrontal cortex (Alexander et al., 1986, 1990; Mink, 1996). The ventral striatum primarily consists of the nucleus accumbens, although ventral putamen and caudate are also considered being part of the ventral striatum (Berns, McClure, Pagnoni, & Montague, 2001; Seger & Cincotta, 2005). Although the dissociation between dorsal-ventral striatum is important to explain cognitive versus more affective aspects of learning, recent studies have suggested that the putamen and caudate of the dorsal striatum similarly may make

contributions to dissociable aspects of action-based learning (Haruno & Kawato, 2006a, b).

In a recent fMRI study, Haruno and Kawato (2006a) studied the activity of putamen and caudate nucleus during the performance of a probabilistic learning task. 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 predicated upon the 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).

A probabilistic learning task measured the evolvement of expectancies about the outcomes of a decision as subjects attempted to learn novel stimulus-action-reward associations. Subjects were instructed to maximize their total monetary reward by pressing a left or right button press when a fractal stimulus appeared on the screen. In order to maximize their reward to a particular stimulus, participants had to discover which response led to a reward most of the time. That is, the outcomes of a response to a stimulus were probabilistically determined, with one response leading to a reward most of the time for a particular stimulus (e.g., 90%, 80%, or 70% of the time rewarded for separate blocks of trials), and the alternative response leading to reward only a small percent of the time (10%, 20%, or 30% for separate blocks of trials). Thus, through trial-and-error, the subject had to develop expectations about the potential reward of a decision until stimulus-action-reward associations were established. Two aspects of learning were distinguished by the authors. First, a reward-prediction error (RPE) value was computed, which was used to infer how proficient subjects were at using errors between anticipated rewards and actual rewards as a basis for adjusting decision-making on future trials. The authors reasoned that this value was particularly meaningful in the early stages of learning when subjects have little experience with which actions maximize rewards and consequently make a higher number of unrewarded decisions. Second, stimulus-action-dependent reward prediction (SADRP) values were computed to capture the proficiency of learning which response maximized reward for each stimulus. Higher SADRP values reflect more effective learning of

stimulus-action-reward associations, and hence, are maximal at the later stages of the task.

The fMRI results 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).

Based on these functional patterns, we used a very similar probabilistic learning task to differentiate the influences of DA medication for the treatment of PD on reward processing and the formation of stimulus-action-reward associations.

Dopamine Modulation of Reward-Based Learning: Evidence from Parkinson's Disease

Several lines of research, including studies of DA-deficient populations, human drug studies, animal studies, and computational modeling, have indicated that DA, via projections from the substantia nigra and ventral tegmental area to the dorsal and ventral striatum, respectively, plays a modulatory role in aspects of reward- and action-

based learning (Arnsten, 1998; Cools, 2006; Daw, Niv, & Dayan, 2005; Eyny & Horvitz, 2003; Frank, 2005; O'Reilly & Frank, 2006; Schultz, 2002). For example, human and primate studies reveal midbrain DA neuronal firing that is timed to reward, especially if it is unexpected (Koepp et al., 1998; Pappata et al., 2002; Schultz, 2002). If a stimulus precedes and reliably predicts the delivery of a reward, the timing of the firing of DA neurons will shift from the reward itself to the onset of the stimulus as learning evolves (Schultz, 2002). This shift in DA firing from reward to antecedents of the reward forms the basis of the temporal-difference learning theory of DA, which states that links between stimuli and responses are adjusted to minimize error between predicted and actual outcomes (the temporal difference error). These prediction errors are coded by changes in firing rate of the DA neurons. These findings provide a strong link between DA function and aspects of reward processing and learning.

Understanding the role of DA in learning is particularly important for individuals who suffer disease or insult that affects the DA system. PD represents one of the more dramatic examples of human DA disease. Studies of PD patients are important from a clinical perspective, but also provide a complementary approach to investigate the role of the basal ganglia and DA function 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). Relevant to the present investigation are findings that the DA projections to regions of the striatum are affected differentially by the progression of PD. PD is initially characterized by DA depletions in the striatum that produce motor deficits, such as tremor, bradykinesia, and rigidity (McAuley, 2003), involving the motor loop (including putamen and supplementary motor areas). Subsequently, these effects extend to the dorsolateral loop (including the DLPFC and the dorsolateral head of the caudate) and still later to the orbitofrontal loop (lateral OFC, ventromedial head of caudate) and the anterior cingulate loop (involving the anterior cingulate cortex and the ventral striatum, in particular the nucleus accumbens) (Kaasinen & Rinne, 2002); these effects are associated with 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).

The primary treatment of PD aims to increase DA availability and activity, including, most prominently, medications functioning as a DA precursor (typically levo-dopa) or as a DA agonist (Hornykiewicz, 1974). However, regions of the striatum are differentially affected by the disease; hence DA medication differentially affects these structures and their related functions. Although DA pharmacotherapy successfully improves motor deficits in PD, its effects on cognitive processes are more controversial. For example, in a critical review of the literature, Cools (2006) concluded that DA medication can have positive and negative consequences on cognitive performance among PD patients. For example, certain cognitive functions, such as task-switching that rely on the heavily DA-depleted dorsolateral and motor loops, improve with DA pharmacotherapy, whereas other aspects of cognition, such as reversal and extinction learning, that depend on ventral circuitries of the basal ganglia and remain relatively spared in early PD, are impaired by DA medication (Cools et al., 2001). These contrasting patterns led to the "overdose" hypothesis, which attempts to account for the negative effects of DA medication on certain cognitive processes (Cools et al., 2001; Czernecki et al., 2002; Gotham, Brown, & Marsden, 1988; Swainson et al., 2000). However, not all aspects of reward-based decision-learning are compromised by DA medication. For example, Shohamy et al., (2005) found that feedback-based learning improved when PD patients were ON DA medication compared to when they were OFF medication.

To account for some of these discrepancies, Frank (2005) emphasized that DA effects on learning depend on the nature of the feedback. Frank argued that a reduction in baseline DA levels, as occurs in untreated PD, disrupts learning based on positive feedback (because the phasic bursts of DA are less effective), while learning based on negative feedback is retained. However, with the addition of DA enhancing medication, learning from positive feedback is restored, but presents an unwanted side effect of overdosing ventral corticostriatal circuits, which impairs learning based on negative feedback (because the phasic dips are ineffective).

Although the mechanism by which these dopaminergic processes are involved in reward-based decision-learning is still unclear, most studies converge on the notion that striatal regions play a key role (Haruno & Kawato, 2006a,b; Knutson et al., 2001; McClure et al., 2003; O'Doherty et al., 2004; Seger & Cincotta, 2005; Tricomi et al., 2004). However, the modulatory role of DA in the caudate versus putamen during

reward-based decision-learning is not yet well established. The current study aims to distinguish the effect of a DA modulation on two structures within the striatum that are involved in reward-based learning; the caudate and putamen.

Present study

The present study investigated the effect of DA modulation on reward-based decision-learning. PD patients performed the previously mentioned probabilistic learning task (Haruno & Kawato, 2006a) both ON and OFF DA medication (within-subjects). We determined the effect of medication on reward-prediction errors during the early phase of learning and on formation of stimulus-action-reward associations during the last phase of learning.

In accordance with patterns of disease progression in PD (Bjorklundt & Dunnett, 2007; Kaasinen & Rinne, 2002), the DA medication should enhance motor-related functions supported by the severely depleted dorsal striatum (in particular the putamen and associated motor circuitry). Therefore we predicted that DA medication would have beneficial effects on the formation of stimulus-action-reward associations. Less pronounced effects were anticipated with respect to RPE values, since the dorsal and especially ventral parts of the caudate are thought to be less depleted from DA compared to the putamen.

Method

Table 1

PD patient information

	Sample N = 20	
	Mean	**SE**
Age (yrs)	68.5	1.64
Gender (male/female)	14/6	
MMSE	28.7	0.26
Years since disease onset	8.08	1.34
L-Dopa (daily dose mg)	542.5	63.27
Finger tapping ON (# taps)	39	2
Finger tapping OFF (# taps)	40	2
Pegboard ON (time in sec)	32	2.3
Pegboard OFF (time in sec)	33	2.0

Participants

Our study included 20 PD patients treated with anti-parkinsonian medication (L-dopa and DA agonist). Summaries of relevant patient details can be found in Table 1. Patients with a Mini-Mental State Examination (MMSE, Folstein, Folstein, & McHugh, 1975) score lower than 25, history of major psychiatric disorders, psychoactive medication, alcoholism, stroke, neurosurgical operation or any other condition known to impair mental status other than PD were excluded. 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.

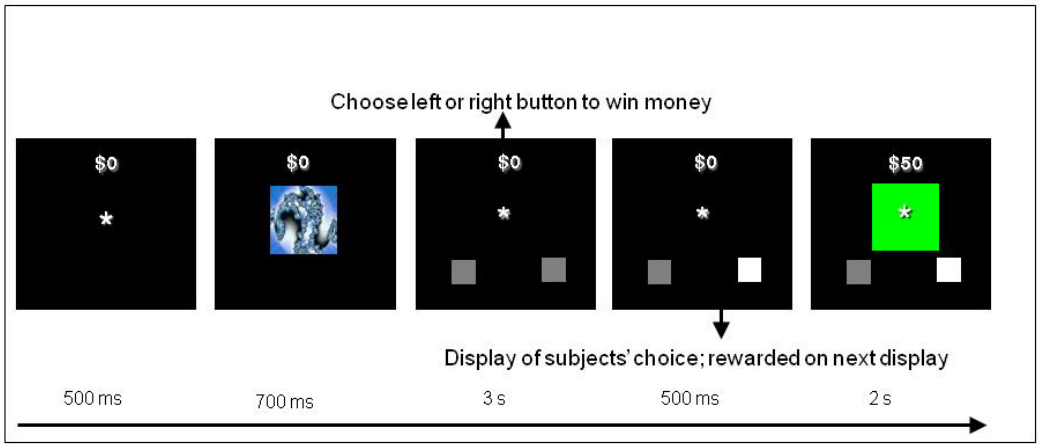
Task and apparatus

Questionnaires

The MMSE (Folstein et al., 1975) assessed the global cognitive state of each patient to verify the absence of dementia. To capture the effects of DA medication 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), was implemented on an IBM-compatible computer with a 17-inch digital display monitor. The computer screen, placed at a distance of ~90 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 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 selected 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. Patients completed the task ON their anti-parkinsonian treatments (L-dopa, DA agonist) and OFF medication. The order of testing with respect to the status of the medication 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 ON and OFF medication. Testing OFF medication took place after a 12h withdrawal period after which L-dopa blood plasma concentrations have been reduced to zero (Crevoisier, Monreal, Metzger, & Nilsen, 2003; Gasser, Jorga, Crevoisier, Hovens, & van Giersbergen, 1999). The order of the ON-OFF sessions was counterbalanced across participants.

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, b; 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 finger tapping test and pegboard was analyzed separately by a one-tailed paired samples t-test. We expected motor performance to improve ON compared to OFF medication. A one-sample t-test was used to test whether MMSE scores (OFF medication) were significantly larger than 25.

The analyses on reward-based decision-learning were based on the mean RPE value (calculated on the first 24 trials) and the mean SADRP value (based on the last 24 trials). First, SADRP and RPE values were separately analyzed by Repeated-Measures ANOVAs, with within-subjects variables Medication (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, Session 90/10 was compared with Session 80/20 and 70/30.

Since individual disease characteristics of PD patients may affect cognitive performance, like disease duration, age and medication dosage (Kaasinen & Rinne, 2002), we will also take these variables into account. Disease duration, daily L-dopa dosage and age were correlated with the dependent variables (improvement in RPE and SADRP comparing ON and OFF medication) 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, age and medication dosage). Note that small RPE values are expected ON medication and high RPE values OFF medication. Thus negative ΔRPE indicates that participants improved, whereas positive ΔRPE indicates that they were impaired ON compared to OFF medication. SADRP values are expected to increase ON versus OFF medication; therefore high ΔSADRP indicates improved performance.

Next, a stepwise regression analysis was performed with the variables that turned out to significantly correlate with RPE and SADRP, that is, L-dopa dosage, disease duration and age with ΔRPE and ΔSADRP in the 90/10 Session. Dependent variables were ΔSADRP in the 90/10 Session and ΔRPE in the 90/10 Session. Independent variables consisted of L-dopa dosage, disease duration and age.

Results

Motor performance

Finger tapping, t (19) = - 0.648, p = 0.50, and pegboard performance, t(19) = -0.19, p = 0.85 were not significantly better ON medication than OFF medication. MMSE scores OFF medication were significantly larger than 25, M = 28.7, t (19) = 14.1, p < 0.001, indicating that our participants were not demented.

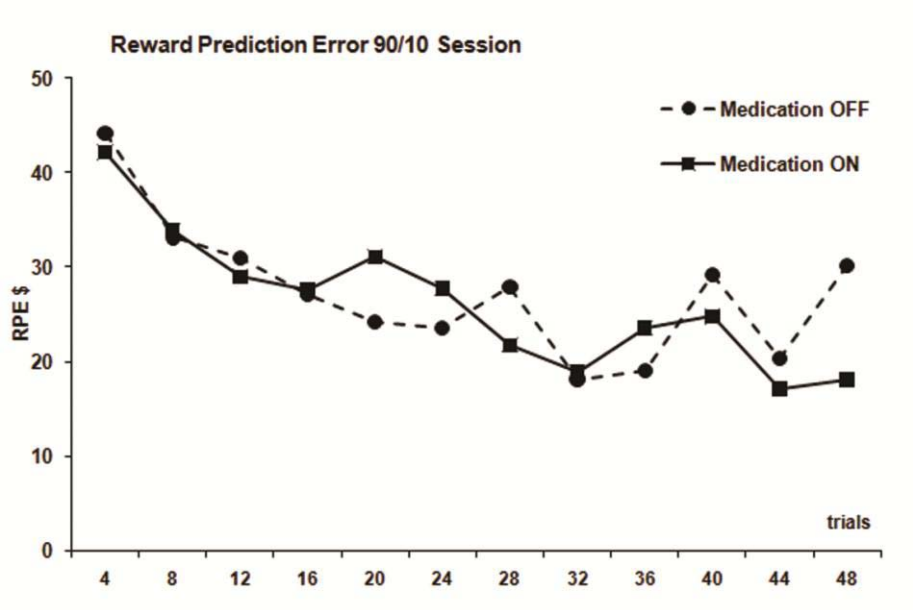
Reward-Based Decision-Learning

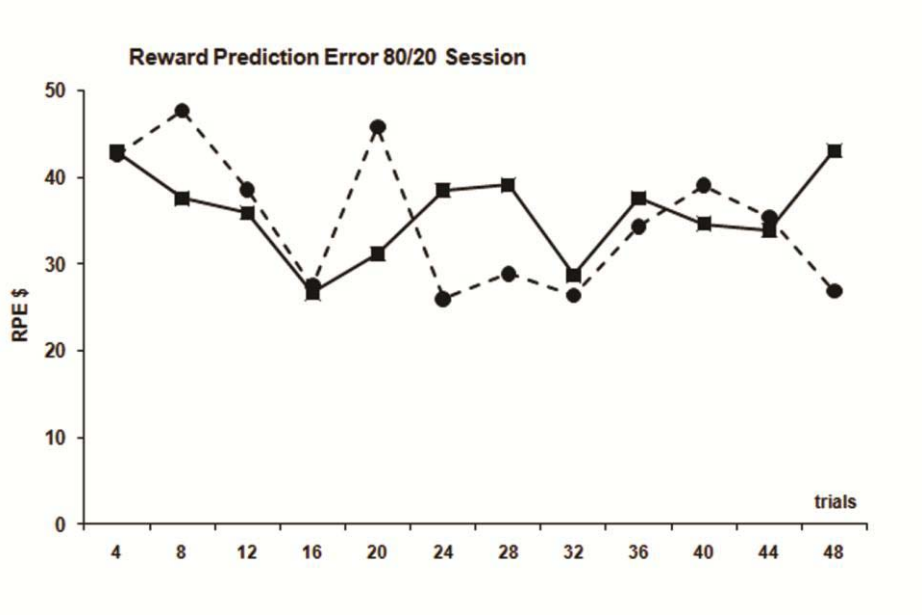
Figure 2 shows the course of mean RPE and SADRP values ON and OFF medication during the task (for each of the sessions). Figure 3 displays the mean RPE values from the first part of each session and the mean SADRP values from the second part of each

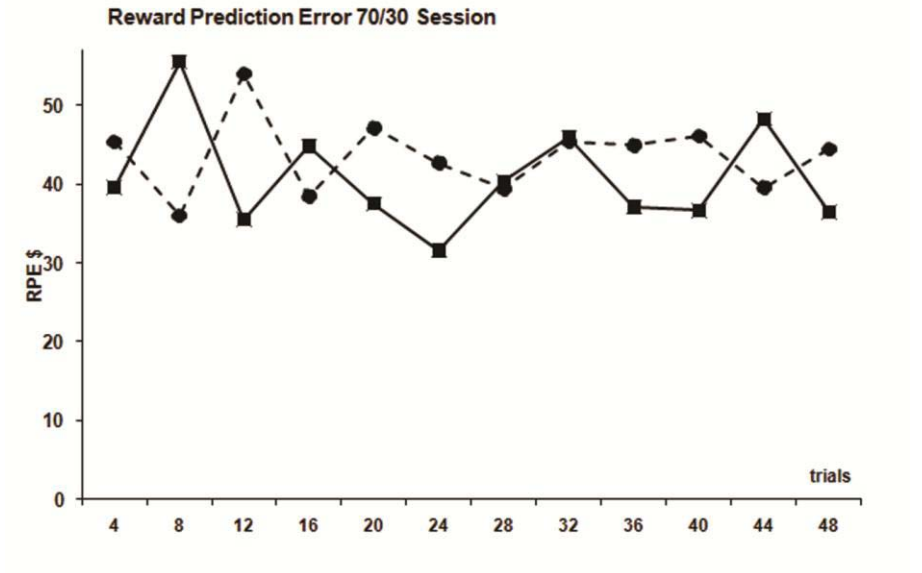
session. Figure 4 shows ΔSADRP and ΔRPE ON-OFF in the 90/10 Session plotted as a function of L-dopa dosage.

Figure 2

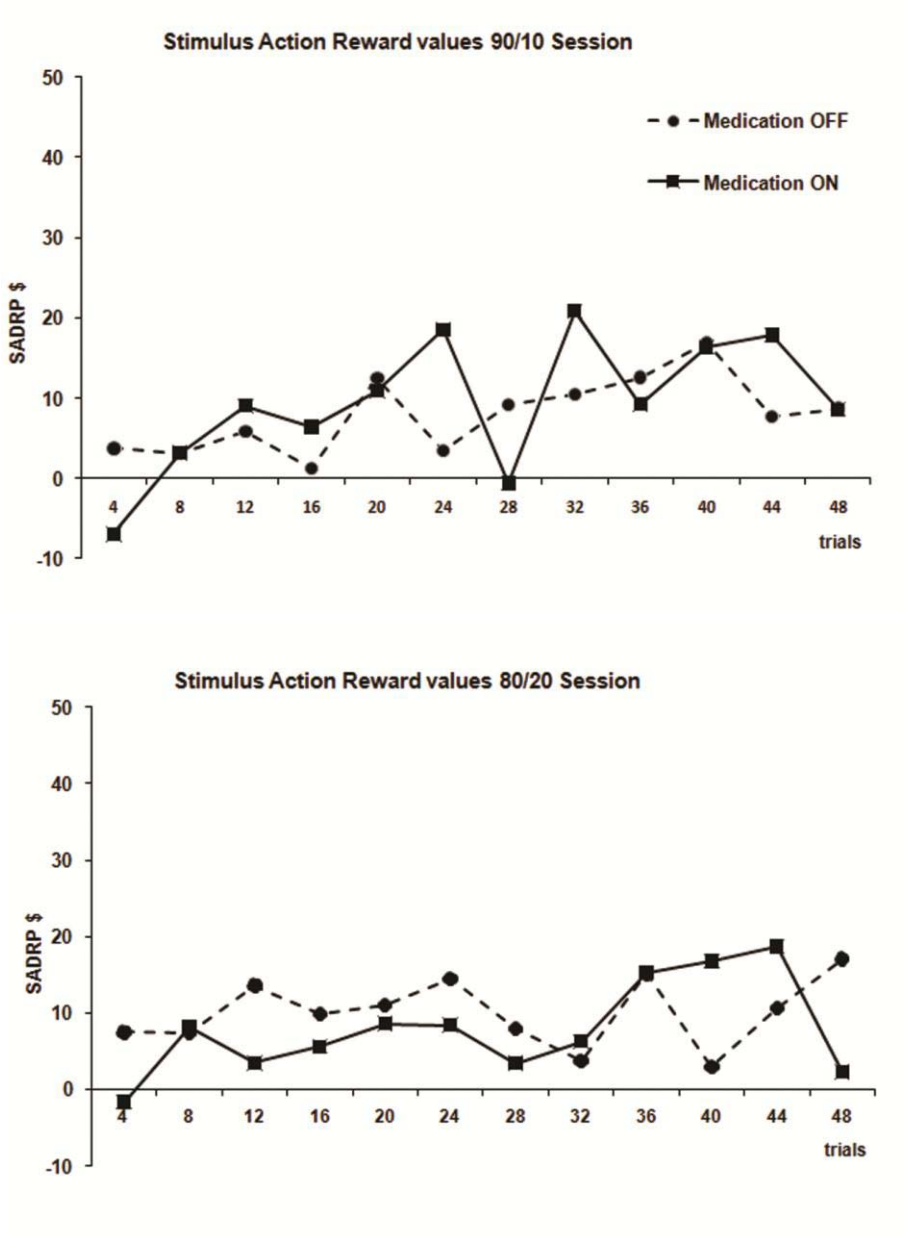
a. Trial-by-trial course of mean RPE values ON and OFF medication, separate for each session.




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


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

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



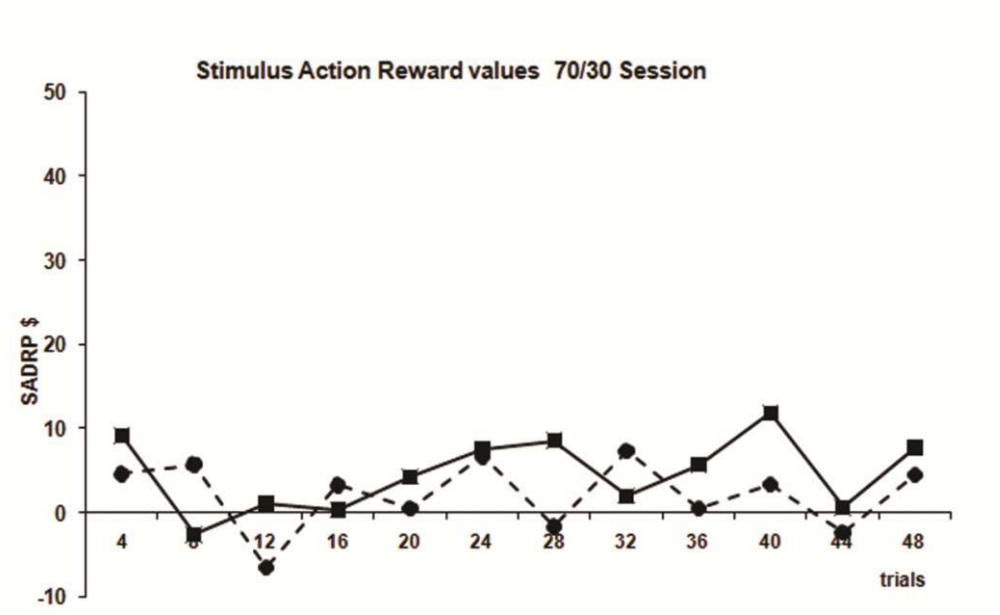
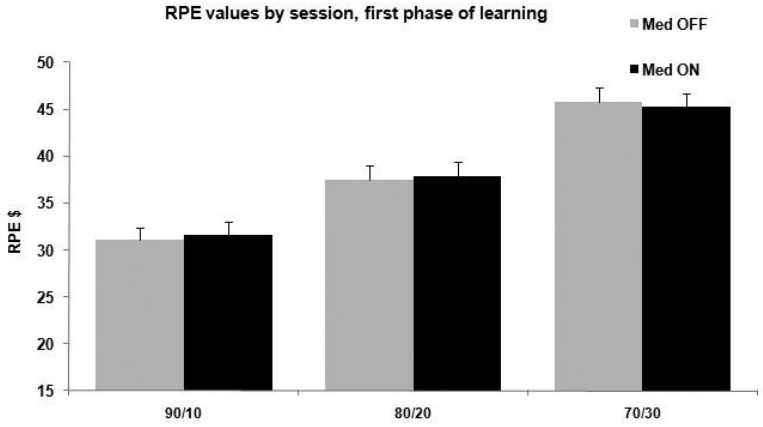
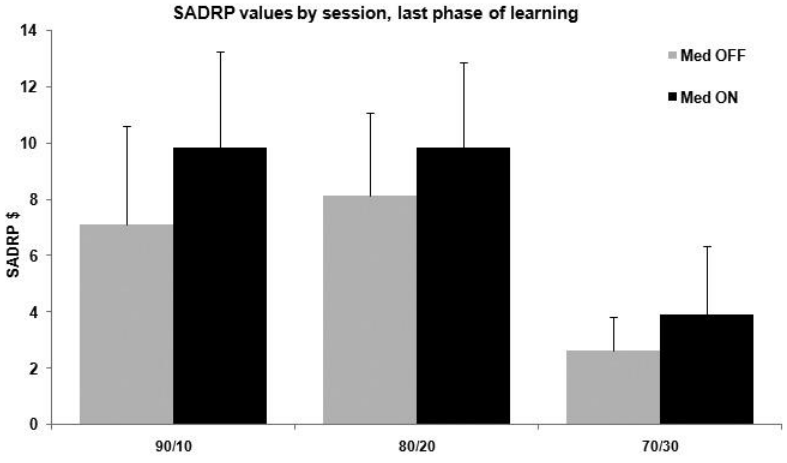

Stimulus Action Reward values 70/30 Session
SADRP $
50
40
30
20
10
0
-10
4
8
12
16
20
24
28
32
36
40
44
48
trials

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.



RPE

The RPE ANOVA revealed a significant effect of Session, F (2, 38) = 36.31, p < 0.0001. RPE values in the 90/10 Session ($M_{90/10}$ = 31.31) were significantly smaller compared to the 80/20 and 70/30 Session, $F_{90/10\text{-}80/20}$ (1, 19) = 17.07, p < 0.01, $M_{80/20}$ = 37.66; $F_{90/10\text{-}70/30}$ (1, 19) = 63.52, p < 0.001, $M_{70/30}$ = 45.55, see Figure 2a and 3a. However, no significant effect of Medication was found, F (1, 19) = 0.01, nor was there an interaction of Medication and Session, F (2, 38) = 0.09.

SADRP

The SADRP ANOVA showed a nearly significant effect of Session, F (2, 38) = 3.10, p = 0.06. SADRP values in the 90/10 session ($M_{90/10}$ = 8.46) were marginally larger compared to the 70/30 Session, $F_{90/10\text{-}70/30}$ (1, 19) = 3.9, p < 0.06, $M_{70/30}$ = 3.27, see Figure 2b and 3b.

The SADRP ANOVA revealed no significant effect of Medication, F (1, 19) = 0.78, nor an interaction of Medication and Session, F (2, 38) = 0.05.

Table 2

ΔSADRP and ΔRPE ON compared to OFF medication correlated with L-dopa dosage, disease duration and age.

Variables	**L-dopa**	**Disease duration**	**Age**
L-dopa (daily mg)	1	0.77**	0.23
Disease duration(yrs)	0.77**	1	0.29
Age (yrs)	0.23	0.29	1
ΔRPE 90/10 session	0.48*	0.20	0.16
ΔRPE 80/20 session	0.04	-0.08	0.01
ΔRPE 70/30 session	0.48*	0.20	-0.20
ΔSADRP 90/10 session	-0.74**	-0.43	0.09
ΔSADRP 80/20 session	0.12	-0.02	-0.02
ΔSADRP 70/30 session	-0.01	0.07	-0.02

*p<0.05 ** p<0.01

Table 3

a. Linear stepwise regression analyses on ΔRPE ON-OFF in the 90/10 Session as a function of L-dopa dosage.

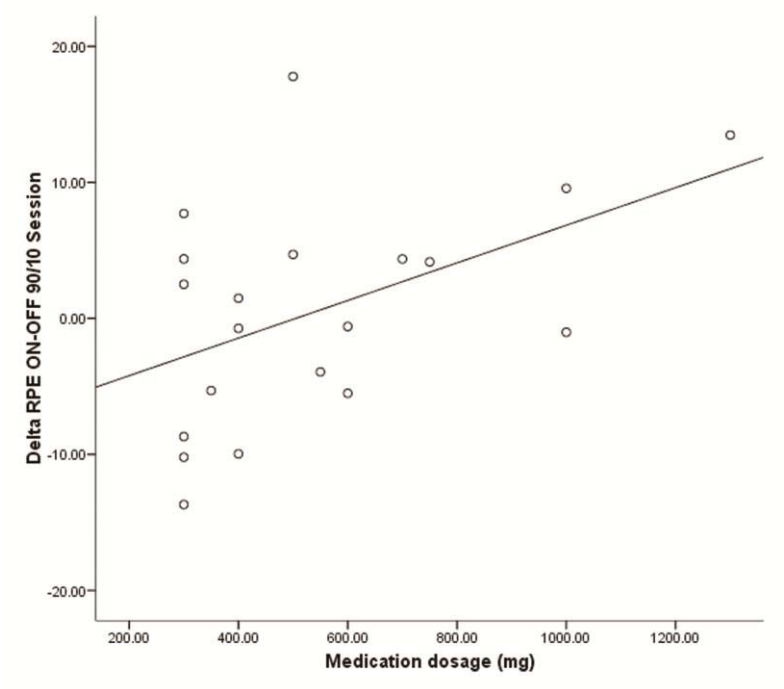
Variables	**β**	**R**	**ΔR^2**
Step 1 L-dopa dosage	0.48	0.48	0.23

b. Linear stepwise regression analyses on ΔSADRP ON-OFF in the 90/10 Session as a function of L-dopa dosage.

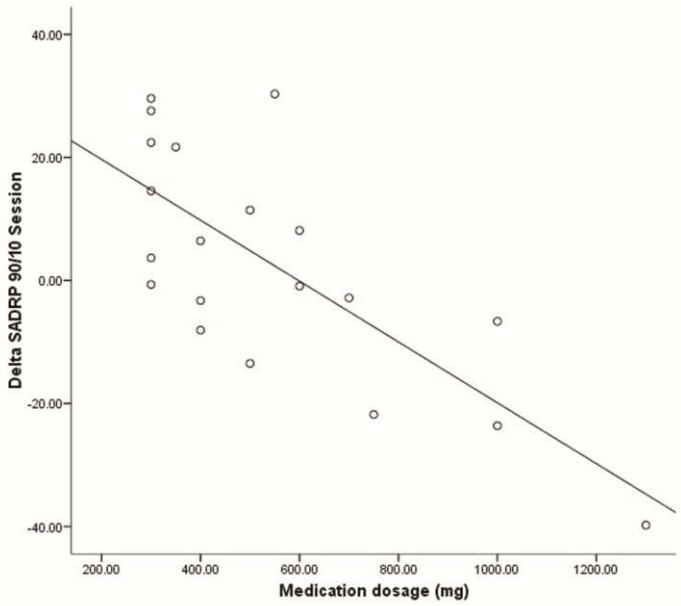
Variables	**β**	**R**	**ΔR^2**
Step 1 L-dopa dosage	-0.74	0.74	0.55

Figure 4

a.ΔRPE ON-OFF in the 90/10 Session as a function of L-dopa dosage.



b. ΔSADRP ON-OFF in the 90/10 Session as a function of L-dopa dosage



Relation ΔSADRP and ΔRPE ON-OFF with daily L-dopa dosage, disease duration and age

Note that small RPE values are expected ON medication and high RPE values OFF medication. Thus high ΔRPE indicates that participants hardly improved or were even impaired ON compared to OFF medication. SADRP values are expected to increase ON versus OFF medication, therefore high ΔSADRP indicate improved performance.

ΔRPE ON-OFF in the 90/10 and 70/30 Session correlated significantly with daily L-dopa dosage, $r_{90/10} = 0.48$, $p < 0.05$; $r_{70/30} = 0.48$, $p < 0.05$. Higher L-dopa dosages are associated with a smaller reduction in RPE values ON compared to OFF medication, that is, patients with a relatively high L-dopa dosage performed worse ON compared to OFF medication early in learning. See Table 2 for correlations.

ΔSADRP ON-OFF in the 90/10 Session revealed a negative correlation with daily L-dopa dosage, $r = -0.74$, $p < 0.001$. The higher the L-dopa dosage, the smaller the increase in SADRP values ON compared to OFF medication. That is, in later stages of learning,

patients with a relatively high L-dopa dosage performed worse ON compared to OFF medication.

Other background variables (disease duration and age) did not significantly correlate with RPE and SADRP values from any of the sessions, see Table 2 for correlations.

Next to that, two stepwise regression analyses were performed. The first analysis included ΔRPE ON compared to OFF (ΔRPE in the 90/10 Session as a dependent variable. The second analysis contained ΔSADRP in the 90/10 Session as a dependent variable. Independent variables for both analyses were entered stepwise: daily L-dopa dosage entered as a first step, disease duration entered as a second step and age as a third step.

Daily L-dopa dosage turned out to be a significant predictor of ΔRPE in the 90/10 Session, F (1, 19) = 5.34, $p < 0.05$. High L-dopa dosages predict a small reduction in ΔRPE ON compared to OFF medication. ΔSADRP in the 90/10 Session was predicted exclusively by L-dopa dosage, F (1, 19) = 22.00, $p < 0.001$. The remaining variables in both analyses did not explain any additional variance and were excluded from the regression model. See Table 3 for R and ΔR^2 for each of the predictive variables and Figure 4 for a scatterplot with ΔRPE ON-OFF and ΔSADRP ON-OFF plotted as a function of L-dopa dosage.

Discussion

The present study aimed to test the effect of a DA modulation on the striatum in probabilistic reward-based decision-learning. First, our results were similar to the behavioral findings of Haruno & Kawato (2006a); performance improved with an increase in predictability of stimulus-action-reward relations.

Second, we predicted that DA medication would improve SADRP values. These predictions were tested by means of a within-subjects design; PD patients performed the probabilistic learning task both ON and OFF medication. DA medication affected the reward-prediction error and stimulus-action-reward-prediction value, functions that have been shown to respectively engage caudate and putamen. During the

experiment, PD patients built up a reward expectation with a specific stimulus-action combination (SADRP). SADRPs, measured at the end of learning, were larger in medicated compared to non-medicated patients, but only in patients with a relatively low medication dosage. SADRP values have been related to activity in the putamen, which suggests that cognitive functions supported by the putamen benefit from the low to moderate amounts of DA medication.

The size of the reward-prediction error early in learning, which is correlated with caudate and ventral striatum activity, was also modulated by DA medication when taking into account medication dosage. The size of the RPE was smaller in patients ON medication compared to OFF medication, although again this pattern was seen exclusively in patients on a low medication dosage. Based on the differential effect of disease progression in PD on caudate and putamen (Bjorklundt & Dunnett, 2007; Kaasinen & Rinne, 2002; Kish et al., 1988), it was predicted that especially SADRP would benefit from DA medication, because putamen is usually more depleted from DA than caudate and ventral striatum early in the disease, thus less pronounced effects were expected for RPE. It turned out that not disease duration, but medication dosage accounts for the effectiveness of medication in reducing the reward-prediction error and strengthening stimulus-action reward associations. How can we explain this effect of medication dosage on both RPE and SADRP?

In a review on DA modulation of cognitive functions in PD patients, Cools (2006) suggested that if performance is impaired in long-range patients ON medication this may be due to earlier and greater L-Dopa doses and fluctuating medication responses rather than greater depletion. With respect to our study, this might explain reduced performance ON medication but not the relatively high performance OFF medication. Studies with different DA polymorphisms have shown contrasting effects of DA drugs on cognitive performance reflecting the genetic variation in baseline levels of DA. Thus individuals with different baseline levels of DA have a different position on the inverted U-shaped curve of optimal performance with DA in PFC (Arnsten, 1998; Goldman-Rakic, Muly, & Williams, 2000). A similar U-shaped curve has been suggested for striatal DA function (Schönberg, Daw, Joel, & O'Doherty, 2007) and this has recently received support. Cools et al. (2009) showed by means of a pharmacological PET study in healthy controls that individual differences in striatal DA synthesis capacity explain positive or negative reward-based learning abilities and differences in

striatal DA drug response. Higher baseline DA synthesis capacity predicts improved reversal learning from positive relative to negative feedback. In response to a D2 agonist, participants with low baseline DA showed enhanced reversal learning from positive relative to negative reward, whereas participants with high baseline DA levels showed a reverse performance pattern. Similarly, DA polymorphisms in PD patients, in addition to their disease duration, may affect their performance-related response to DA medication. That is, early stage PD patients with a low baseline DA may benefit more from DA medication during reversal learning because the striatal areas that are less affected by DA depletion, such as the caudate, will not be overdosed. Future studies should consider measuring these individual differences.

Along the lines of the overdose hypothesis (Cools et al., 2001; Gotham et al., 1988), functions known to rely on the dorsal striatum or dorsolateral loop, such as task-switching, are enhanced with medication (Cools et al., 2001; Gotham et al., 1988), while the ventral circuitry and cognitive functions that rely on this loop are overdosed and impaired. However, we did not find an impairment of RPE (relying on dorsal and ventral caudate) in patients ON medication. The overdose hypothesis mainly explains impaired performance found in reversal and extinction learning (Cools et al., 2001; Czernecki et al., 2002; Swainson et al., 2000). Frank's (2005) modeling work elaborated on this idea and showed that PD patients OFF medication more effectively process negative feedback in comparison to positive feedback whereas PD patients ON medication show the opposite pattern. In the current task a reward-prediction error results from either unexpected positive or negative feedback, thus no difference in RPE or SADRP values between PD patients ON and OFF medication would be observed. However, the current study had a different approach by taking into account the role of the caudate and putamen with respect to reward-based decision-learning and the individual differences in DA medication. Our results provided some insight into the effect of DA medication on caudate and putamen and the associated reward-based decision-learning processes, i.e. reward expectation and building associations between stimulus-action and reward. Both cognitive functions improve with DA medication, although the beneficial effect depends on the amount of medication. It remains an open question whether positive and negative feedback would differentially affect caudate and putamen.

Limitations

There are some limitations related to the experimental paradigm and thereby the interpretation of the results. Although SADRP and RPE 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 yields an increase in SADRP values (according to the computational model). Thus, a null result of medication 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 DA. Rather, it suggests that the medication 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 DA medication 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 learning in PD by means of an fMRI study.

In the 80/20 and 70/30 Session performance was not affected by DA medication. When the chance to receive reward is less predictable, the match between stimulus and response categories becomes less clear, which negatively affects implicit learning (Maddox & Ashby, 2004). With increased probabilistic difficulty, performance may have shifted from an implicit learning strategy to a more explicit rule based learning strategy (Maddox & Ashby, 2004). Rule based performance relies on frontal and medial temporal lobes and may be less affected by DA changes in the BG compared to the implicit learning system.

Conclusion

In sum, both aspects of reward-based learning, the evaluative component (RPE) and action-reward association learning component (SADRP) were affected by DA medication which suggests that their underlying neural structures, caudate and putamen, both benefit from DA modulations in PD patients. However, there seems to be an optimal level of DA medication; i.e. large daily doses may become suboptimal.