Published in final edited form as:

Neuropsychologia. 2012 April ; 50(5): 583–591. doi:10.1016/j.neuropsychologia.2011.12.012.

Dose dependent dopaminergic modulation of reward-based learning in Parkinson's disease

N.C. van Wouwe^{a,b,c,*}, K.R. Ridderinkhof^{d,e}, G.P.H. Band^b, W.P.M. van den Wildenberg^d, and S.A. Wylie^c

^aTNO Defence, Security and Safety, The Netherlands ^bLeiden Institute for Brain and Cognition, Psychology Department, Leiden University, Leiden, The Netherlands ^cNeurology Department, University of Virginia Health Systems, Charlottesville, VA, USA ^dAmsterdam Center for the Study of Adaptive Control in Brain and Behaviour (ACACIA), Psychology Department, University of Amsterdam, Amsterdam, The Netherlands ^eCognitive Science Center Amsterdam, University of Amsterdam, Amsterdam, The Netherlands

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 with actions that lead to rewarding outcomes. Mathematical models of reinforcement-based learning to select rewarding actions distinguish between (1) the formation of stimulus–action–reward associations, such that, at the instant a specific stimulus is presented, it activates a specific action, based on the expectation that that particular action will likely incur reward (or avoid punishment); and (2) the comparison of predicted and actual outcomes to determine whether the specific stimulus–action association yielded the intended outcome or needs revision. Animal electrophysiology and human fMRI studies converge on the notion that dissociable neural circuitries centered on the striatum are differentially involved in different components of this learning process. The modulatory role of dopamine (DA) in these respective circuits and component processes is of particular relevance to the study of reward-based learning in patients diagnosed with Parkinson's disease (PD). Here we show that the first component process, learning to predict which actions yield reward (supported by the anterior putamen and associated motor circuitry) is impaired when PD patients are taken off their DA medication, whereas DA medication has no systematic effects on the second processes, outcome evaluation (supported by caudate and ventral striatum and associated frontal circuitries). However, the effects of DA medication on these processes depend on dosage, with larger daily doses leading to a decrease in predictability of stimulus–action–reward relations and increase in reward-prediction errors.

Keywords

Parkinson's disease; Dopaminergic medication; Caudate; Putamen; Probabilistic learning

1. Introduction

Learning to select behavior to maximize reward in a given situation 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 cortex (PFC) 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 diagnosed with Parkinson's disease (PD) both ON and OFF their DA medication on a probabilistic learning task.

1.1. 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 reward-based decision-learning (Brown & Marsden, 1998; Kimura, 1995; Knowlton, Mangels, & Squire, 1996; Packard & Knowlton, 2002; Schultz, Tremblay, & Hollerman, 2003).

Lesion and human imaging studies demonstrate an important contribution of the striatum to reward-based decision-learning and support a functional dissociation between dorsal and ventral areas of the striatum (for an overview see Balleine, Delgado, & Hikosaka, 2007). Dorsal portions of the striatum are 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 depletion in these areas disrupt formation of stimulus–response associations (Faure, Haberland, Condé, & El Massioui, 2005; Yin, Knowlton, & Balleine, 2004). Activation of 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 or RPE; Knutson, Fong, & Hommer, 2001; McClure, Berns, & Montague, 2003; O'Doherty et al., 2004; Seger & Cincotta, 2005).

In addition to this functional dissociation, recent imaging studies have suggested that the putamen and caudate of the dorsal striatum may contribute to dissociable aspects of action-based learning. For instance, Haruno and Kawato (2006a) studied BG activity in healthy participants during performance on a probabilistic learning task, in which participants had to discover for each particular stimulus whether a right or a left button press led to a reward most of the time; through trial-and-error, they learned stimulus–action–reward associations. A *Q*-learning model was used to generate individual parameters that reflect two important aspects of learning. First, the mismatch between anticipated rewards and actual rewards was computed as a reward-prediction error (RPE), which learners used for adjusting decision-making on future trials, in particular in the early stages of learning when they relied on feedback to determine which actions maximize rewards. Higher RPE values were associated with activation of the caudate nucleus and ventral striatum and their associated frontal circuitry (orbitofrontal, dorsolateral prefrontal, and anterior cingulate cortex), involved in generating and testing hypotheses regarding reward optimization (cf. Alexander et al., 1990; Oyama, Hernadi, Iijima, & Tsutsui, 2010). Second, as learning progressed, participants should have been able to forecast which actions would likely yield reward (or avoid punishment); this was computed as the stimulus–action-dependent reward prediction (SADRP). Higher SADRP values reflected more effective learning of stimulus–action–

reward associations, and hence, were maximal at the later stages of the task. Higher SADRP values were associated with activation of the anterior putamen and its associated motor circuitry (supplementary motor area, premotor and primary motor cortex), involved in integrating information on the expectation of reward with processes that mediate the actions leading to the reward (cf. Alexander et al., 1990; Gerardin et al., 2003). Thus, during later stages of learning, putamen activity increased with reward predictions (i.e., with learning SADRP).

Activity in the putamen increased 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 became more predictable and learning was gradually fine-tuned (Haruno & Kawato, 2006b). As these SADRP values increased, the RPE was reduced as subjects more accurately anticipated the rewards associated with their actions. The authors argued that the global reward-related features of these stimulus–action–reward associations appeared to propagate from the caudate to motor loops (which include the putamen and premotor areas), likely by means of a DA signal subserved by reciprocal projections between the striatum and the substantia nigra (Haruno & Kawato, 2006b). Interestingly, this 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).

1.2. Dopamine modulation of reward-based learning

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; Eyni & Horvitz, 2003; Frank, 2005; O'Reilly and 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 cue 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 (i.e., 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.

1.3. Reward-based learning in Parkinson's disease

Understanding the role of DA in learning is particularly important when considering neurological conditions that disrupt the DA system. PD represents one of the more dramatic examples of human DA dysfunction that results in marked changes in motor and cognitive functioning. 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; Bjorklund & Dunnett, 2007).

The primary treatment to reduce PD motor symptoms, which commonly include tremor, bradykinesia, and rigidity, aims to increase DA availability and activity, including, most prominently, medication functioning as a DA precursor (typically levodopa) or as a DA agonist (Hornykiewicz, 1974). Because regions of the striatum are differentially affected by the disease, 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 equivocal. 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, Barker, Sahakian, & Robbins, 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; Swinson et al., 2000).

However, not all aspects of reward-based decision-learning are compromised by DA medication. For example, Shohamy, Myers, Grossman, Sage, and Gluck (2005) found that feedback-based learning improved when PD patients were ON DA medication compared to when they were OFF medication. Frank, Seeberger, and O'Reilly (2004) showed that this benefit obtained specifically for learning that certain actions are likely to yield reward, whereas learning that certain other actions are likely to yield punishment was negatively affected by DA medication. This pattern of levodopa-induced improved incentive learning but impaired avoidance learning, replicated by Bódi et al. (2009), is taken to reflect strengthened disinhibition along the direct route and weakened inhibition along the indirect route within the basal ganglia.

Although studies converge on the notion that striatal regions play a key role in reward-based decision-learning (Bódi et al., 2009; Cools et al., 2009; Frank et al., 2004; Haruno & Kawato, 2006a, 2006b; Knutson et al., 2001; McClure et al., 2003; O'Doherty et al., 2004; Seger & Cincotta, 2005; Tricomi et al., 2004), the modulatory role of DA in different structures within the striatum is not yet well established. DA might have dissociable effects on different component processes of reward-based decision-learning, for example, on outcome evaluation processing supported by caudate and ventral striatum or on reward prediction processing supported by anterior putamen.

1.4. Present study

The present study investigates 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 (RPE) during the early phase of learning and on formation of stimulus–action–reward associations (SADRP) during the last phase of learning.

In accordance with patterns of disease progression in PD (Bjorklund & Dunnett, 2007; Kaasinen & Rinne, 2002), 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.

2. Methods

2.1. Participants

Our study included 20 PD patients (6 females; mean age, 68.5 years) treated with anti-parkinsonian medication (L-dopa and D2 agonist). Eight patients were on DA agonists (pramipexole or ropinirole) in addition to L-dopa, whereas the remaining patients were exclusively treated with L-dopa. Dopamine precursors and dopamine agonists were converted into one value representing L-dopa equivalent daily doses (LEDD) values (Weintraub et al., 2006). Summaries of relevant patient details can be found in Table 1. Patients with a 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.

2.2. Task and apparatus

2.2.1. 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 DA medication on fine motor dexterity and speed, we administered the Purdue Pegboard task (Lezak, 1995) and a finger-tapping test during each condition. 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 s. The tapping task alternated between each hand until three attempts were completed with each hand.

2.3. 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 RPE 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 in game money (note: participants were not remunerated for their participation). Fig. 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 $5.67^\circ \times 4.41^\circ$ (9 cm $\times$ 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 Fig. 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 s 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 conditions of 48 trials. For each condition, 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 non-optimal choice; (2) in the first condition, selecting the optimal response hand resulted in a 90% probability of winning \$50 and a 10% probability of losing \$50; (3) in a second condition, selecting the optimal response hand resulted in an 80% probability of winning \$50 and a 20% probability of losing \$50; (4) in a third condition, selecting the optimal response hand resulted in a 70% probability of winning \$50 and a 30% probability of losing \$50. In all conditions, the probabilities of winning versus losing were reversed for the non-optimal relative to the optimal response hand. As an example, in the 90/10 condition a left response to fractal stimulus 1 (FS1) yielded a \$50 reward with a probability of 0.9 (90%) and a \$50 loss with a probability of 0.1 (10%). A right response to FS1 yielded a \$50 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 condition 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 condition were also 0.9. The optimal response for each fractal was pseudorandomized 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 condition, the specific response options were randomly attached to each of the fractals. Additionally, the fractal stimuli were presented randomly and with equal frequency within a condition. Condition order was also randomized.

2.4. Procedure

Participants completed two versions of task on different days. 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 12 h withdrawal period after which L-dopa blood plasma concentrations are reduced to zero (Crevoisier, Monreal, Metzger, & Nilsen, 2003; Gasser, Jorga, Crevoisier, Hovens, & van Giersbergen, 1999). Note however that DA agonists are associated with a longer half-life, which may have resulted in some residual medication effects.

2.5. Computational model for estimating SADRP and RPE

A reinforcement model (*Q*-learning, Sutton & Barto, 1998) was used to estimate individual 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 (SADRP) for each action (two responses) and each fractal stimuli (three different fractal stimuli) will be calculated at time t , $Q_t(\text{FS}, r)$ which adds up to six SADRPs values per condition. The RPE represents the actual reward received (rt) minus the expected reward, $\text{RPE} = rt - (Q_t(\text{FS}, r))$. For the next occurrence of the same stimulus and action, SADRPs and RPE values are updated according to the “ Q -learning algorithm” to maximize reward (Sutton & Barto, 1998), $Q_{t+1}(\text{FS}, r) = Q_t(\text{FS}, r) + a_t^{\text{FS}} (rt - (Q_t(\text{FS}, r)))$. The initial value of the learning rate for the fractal stimuli is ‘1’ and this value is equal for all subjects.

The learning rate is updated separately for each FS according to the following rule:

$a_t^{\text{FS}} = (a_{t-1}^{\text{FS}}) / (1 + a_{t-1}^{\text{FS}})$. 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 SADRPs become reliable). This is an important feature of a_t^{FS} because it means that, at the end of learning, the SADRPs are 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 as in Haruno & Kawato, 2006a), while the SADRPs value is small. Major changes in SADRPs are especially expected in the first stage of learning. In a later stage of learning (i.e., last 24 trials) SADRPs become accurate and does not show large changes (converges to an asymptotic value). Additionally, RPEs are expected to be small at the end of learning.

2.6. 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.

For the probabilistic reward learning task, we first calculated the cumulative accuracy (cumulative percentage correct over trials) for each condition as a function of medication to investigate whether medication would affect general learning over time. Cumulative accuracy (cumulated accuracy values after the first 24 trials and accuracy after 48 trials) were analyzed by a repeated-measures analysis of variance (RM-ANOVA) including the within subject variables Medication (OFF, ON), Time (First, Second Half) and Condition (90/10, 80/20, 70/30).

Second, SADRPs and RPE values were analyzed separately by repeated-measures analysis of variance (RM-ANOVA), including the within-subjects variables of Medication (OFF, ON) and Time (First, Second Half) and Condition (90/10, 80/20, 70/30) to investigate whether the patients show RPE and SADRPs learning from the first to the second half of the experiment. The analyses were based on the mean RPE value from the first half of the task (calculated on the first 24 trials) and the second half of the task (based on the second 24 trials) and the mean SADRPs value from the first and second half of the task.

Subsequently, SADRPs and RPE values of the theoretically relevant learning phases were analyzed more elaborately. The RPE analyses were based on the mean RPE value calculated on the first half, and the SADRPs analyses on the mean SADRPs value based on the last half

of the experiment. SADRP and RPE values were separately analyzed by Repeated-Measures ANOVAs, with within-subjects variables Medication (OFF, ON) and Condition (90/10, 80/20, 70/30). Condition types are represented as the dominant versus nondominant probability. Specific predictions were tested by using a Simple Contrast test, that is, Condition 90/10 was compared with Condition 80/20 and 70/30.

Since individual disease characteristics of PD patients, like disease duration, age, and medication dosage may affect cognitive performance (Kaasinen & Rinne, 2002), disease duration, LEDD and age were correlated with the dependent variables (change in RPE and SADRP comparing ON and OFF medication and RPE and SADRP OFF medication).

First, we correlated change in RPE (RPE ON minus OFF = Δ RPE) and change in SADRP (SADRP ON minus OFF = Δ SADRP), separately for each condition, 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 Δ RPE indicates that participants improved, whereas positive Δ RPE indicates that they were impaired ON compared to OFF medication. SADRP values are expected to increase ON versus OFF medication; therefore high Δ SADRP indicates improved performance.

Correlations between disease duration and RPE and SADRP OFF medication may provide some additional information regarding effects of disease severity separate from dosage effects.

3. Results

3.1. 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 the participants were not demented.

3.2. Reward-based decision-learning

Fig. 2 presents the cumulative accuracy values by Medication and Condition. Fig. 3 shows the mean RPE values ON and OFF medication separately for the first and second half of the experiment. Fig. 4 displays the mean SADRP values from the first and second part of the experiment. Fig. 5 shows Δ SADRP and Δ RPE ON–OFF in the 90/10 Condition plotted as a function of LEDD.

3.2.1. Cumulative accuracy—Medication produced no significant effect on cumulative accuracy, $F < 1$. Cumulative accuracy increased over Time, ($F(1, 19) = 891.39$, $p < 0.001$) and differed across Conditions ($F(2, 38) = 5.68$, $p < 0.01$) showing higher accuracy values in the 90/10 ($M_{90/10} = 58.70\%$) compared to the 80/20 ($M_{80/20} = 58.60\%$) and 70/30 condition ($M_{70/30} = 52.71\%$) at the end of learning. The Condition effect did not interact with Medication, $F < 1$, or Time, $F(2, 38) = 1.74$, $p = 0.19$. See Fig. 2 for cumulative accuracy values plotted trial-by-trial separate for each Condition.

3.2.2. RPE—RPE values were significantly larger at the beginning of the experiment compared to the end of the experiment, $F(1, 19) = 137$, $p < 0.001$, which indicates that the patients reduced their reward prediction errors over time, see Fig. 3. Additionally, RPE values varied across Conditions, $F(1, 19) = 105.34$, $p < 0.001$, revealing larger RPE values in the 70/30 and 80/20 compared to the 90/10 condition. The Condition effect interacted with Time, $F(2, 38) = 7.44$, $p < 0.01$; the 90/10 condition showed a larger reduction in RPE

from beginning to end of learning compared to the other conditions. Medication did not influence RPE, $F < 1$, nor did Medication interact with Time or Condition, $F < 1$.

Small RPE values are expected ON medication and high RPE values OFF medication. Thus, negative Δ RPE indicates that patients improved whereas positive Δ RPE indicates that patients were impaired when ON compared to OFF medication. Δ RPE in the 90/10 Condition correlated significantly with LEDD, $r_{90/10} = 0.48$, $p < 0.05$. At low levels of LEDD, outcome evaluation was not affected systematically by medication ON versus OFF, whereas higher LEDD was associated with less effective outcome evaluation when ON medication compared to OFF medication (see Table 2 and Fig. 5 for correlations). Other background variables (disease duration and age) did not significantly correlate with Δ RPE values in any of the conditions (see Table 2). RPE OFF medication did not correlate with any of the background variables.

3.2.3. SADRP—The second half of the experiment yielded larger SADRP values than the first half, although the Time effect was not significant, $F(1, 19) = 1.71$, $p = 0.2$, see Fig. 4. SADRP values differed across Conditions, $F(1, 19) = 3.11$, $p = 0.06$, showing significantly larger SADRP values in the 90/10 compared to the 70/30 condition, $F(1, 19) = 4.44$, $p < 0.05$. Overall, Medication did not affect SADRP values, $F < 1$. However, the Medication effect interacted with Time, $F(2, 38) = 6.53$, $p < 0.05$; patients ON medication showed an increase in SADRP from beginning to end of learning trials, whereas very little increase in SADRP occurred from beginning to end OFF medication. Medication did not interact with Condition, $F < 1$, nor did Time interact with Condition $F < 1$.

High SADRP values are expected ON medication and low SADRP values OFF medication. Thus, high Δ SADRP scores indicate that participants improved ON compared to OFF medication. Δ SADRP in the 90/10 Condition revealed a highly significant negative correlation with LEDD, $r_{90/10} = -0.66$, $p < 0.01$. Patients with lower levels of LEDD dosage showed improvements in action-outcome learning when ON compared to OFF medication; for patients with higher daily dosages the net effect of medication ON versus OFF was null or even negative. Other background variables did not significantly correlate with Δ SADRP values in any of the conditions, see Table 2 for correlations.

SADRP OFF medication did not correlate with disease duration or medication dosage. The absence of a correlation between SADRP and RPE OFF medication and disease duration implicates that differences in baseline performance are probably not due to disease severity.

4. Discussion

The present study aimed to test the effect of DA modulation on probabilistic reward-based decision-learning that relies on striatal functioning. First, our results replicate the behavioral findings reported by Haruno and Kawato (2006a); performance improved as a function of time with an increase in predictability of stimulus–action–reward relations.

Second, we predicted that DA medication would improve SADRP values (especially towards the later stages of learning), whereas it would have less pronounced effects on RPE. These predictions were tested by means of a within-subjects design; PD patients performed the probabilistic learning task both ON and OFF medication. Early on during learning, patients relied on feedback to figure out which actions yielded reward. This process of learning by outcome evaluation, reflected in the resulting RPEs, was not influenced by DA medication. If anything, higher medication dosages impaired outcome evaluation; low to moderate dosages yielded no systematic effect. This suggests that the caudate and ventral striatum (and their associated circuitry involved in hypothesis generation and value

updating) do not benefit from DA medication, and in fact, may be detrimentally overdosed at higher doses of DA. As learning progressed, however, PD patients began to build up expectations that specific stimulus–action combination would likely yield reward. These SADRP values were larger when patients were ON compared to OFF medication, suggesting that the anterior putamen (and its associated sensorimotor circuitry involved in action selection and stimulus–action learning) benefits from DA medication. However, this reward prediction (increase in SADRP) benefit was observed only in patients with a relatively low to moderate medication dosages.

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 DL-PFC 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; Swanson et al., 2000; Wylie et al., 2009a, 2009b). Based on the differential effect of disease progression in PD on caudate and putamen (Bjorklund & Dunnett, 2007; Kaasinen & Rinne, 2002; Kish, Shannak, & Hornykiewicz, 1988), especially SADRP would benefit from DA medication, because putamen is usually more depleted from DA than caudate and ventral striatum early in the disease. 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 SADRP and, to a lesser extent, RPE?

In a review on DA modulation of cognitive functions in PD patients, Cools (2006) suggested that performance might not necessarily be impaired by DA depletion. When performance is impaired in patients who are ON medication for several years, this may be due to earlier and greater L-Dopa doses and fluctuating medication responses. With respect to our study, this might explain reduced performance ON medication but not the relatively high performance OFF medication in patients on higher doses of 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 occupy 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). Using PET imaging in healthy adults, Cools et al. (2009) showed that individual differences in striatal DA synthesis capacity predicts positive or negative reward-based learning abilities as well as changes in learning in response to DA drug challenge. Higher baseline DA synthesis capacity was associated with better reversal learning from positive relative to negative feedback compared to individuals with lower synthesis capacity. When administered a D2 agonist, participants with low baseline DA synthesis developed enhanced learning from positive relative to negative reward, whereas participants with high baseline DA synthesis showed a pattern that reversed from their baseline, i.e., learning was enhanced for negative reward relative to positive reward. 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 who have low baseline DA levels to begin with may show improved performance with administration of low to moderate DA medication doses and be resistant to an “overdosing” response in relatively

intact striatal areas until much higher medication dosages are administered. Clearly, future studies are needed to better account for individual variability in baseline performance, baseline dopamine characteristics, and response to treatment.

According to the overdose hypothesis (Cools et al., 2001; Gotham et al., 1988), functions known to rely on the relatively DA depleted dorsal striatum or dorsolateral loop, such as task-switching, are enhanced with medication (Cools et al., 2001; Gotham et al., 1988), while the relatively DA intact ventral circuitry and associated cognitive functions are overdosed and impaired. However, we observed medication-driven impairments of RPE (relying on dorsal and ventral caudate) only in patients who used high daily dosages of DA medication. The overdose hypothesis mainly explains impaired performance found in reversal and extinction learning (Cools et al., 2001; Czernecki et al., 2002; Swinson et al., 2000; Voon et al., 2010). 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 though, a reward-prediction error results from either unexpected positive or negative feedback, thus a preference for positive or negative feedback cannot be distinguished based on SADRP and RPE values.

The current study contributed insights beyond those reported above by focusing on component processes of reward-based decision-learning that rely on different striatal circuits, and by examining individual differences in the effects of DA medication. Our results allow us to articulate with greater precision the effect of DA medication on the caudate and ventral striatum on the one hand and on the putamen on the other. While DA medication leaves outcome evaluation processes (supported by caudate and ventral striatum) seemingly unaffected, or suggestively worsened at high dosages, learning to predict which actions yield reward (supported by the anterior putamen and associated motor circuitry) is improved by DA medication (at low to moderate dosages). Whether positive and negative feedback differentially affect caudate and putamen functioning remains an open question.

4.1. Limitations

There are some limitations related to the experimental paradigm that could affect 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 the early phases of learning.

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). The decay component of the learning rate was defined according to a standard formula that was kept constant for all patients and conditions to be able to specifically investigate changes in the outcome parameters, i.e., SADRP and RPE. Although we did not model the decay of the learning rate, the model-based studies that have estimated learning parameters in PD patients (Rutledge et al., 2009; Voon et al., 2010) seem to be in line with the present study. Rutledge et al. modeled the learning rate in a task with constantly changing reward probabilities. They found higher learning rates in PD patients ON medication compared to OFF medication. Thus, the application of a higher learning rate would have amplified our findings (i.e., smaller RPEs and larger SADRP ON medication). Similar to our findings, Voon et al. (2010) showed that exposure to a larger daily dosage of

DA medication in PD patients leads to a lower percentage of correct responses and a slower learning rate in a reward learning task, although this finding was specific for learning from losses. Inclusion of fMRI measures as a means for tracking the differential involvement and potentially temporal engagement of caudate and putamen during probabilistic reward-based learning in PD would be very informative.

In the 80/20 and 70/30 Conditions, performance was not affected by DA medication. When 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 learning difficulty, performance may have shifted from an implicit learning strategy to a more explicit rule-based learning strategy (Maddox & Ashby, 2004), for example 'every blue stimulus asks for a left response'. 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.

Simple motor functions did not improve with DA medication in the current study, although DA medication was supposed to enhance motor processing through its effect on the cortical "motor" loop. Similar findings have recently been reported by Graef et al. (2010). They showed a null-finding on motor skills together with an improvement in instrumental learning. The results from the finger tapping task and the pegboard test in our study indicate that there were no significant improvements of peripheral motor skills. More pronounced dopaminergic-related improvements on peripheral motor skills might however be expected in PD samples associated with more advanced stages of the disease. Importantly, we obtained DA medication effects on reward learning in the absence of motor effects which suggest that the observed dosage-dependent effects on reward-based learning cannot be explained in terms of, and go beyond, medication-related changes in peripheral motor skills. Moreover, this suggests that changes in the ability to issue a manual response ON and OFF medication do not explain the difference in performance.

Additionally, there was no effect of disease duration on reward-based learning. However, Rowe et al. (2008) showed reduced ACC activation with reward expectation in patients with more advanced disease duration as well as an impaired behavioral response to reward signaling cues. Thus, more sensitive measures indicative for the degree of cortical degeneration with disease duration, like imaging data, might provide further insight into the interaction between disease duration and medication effects on reward-based learning processes.

4.2. Conclusion

In sum, the present findings highlight the effect of DA medication on two aspects of reward-based learning, the evaluative component (RPE) and the action-reward association learning component (SADRP). While DA medication leaves evaluative aspects of new learning supported by caudate and ventral striatum largely unaffected, or even worsened at high dosages, DA medication, particularly low and moderate doses, improves the action-reward learning processes supported by the anterior putamen and associated motor circuitry. Notably, evaluative and action-reward association components of learning were compromised by high doses of DA medication.

Acknowledgments

NCvW was supported by a travel grant from "Leids Universitair Fonds" (LUF). The work of KRR, GPHB and WPMvdW was supported by grants from the Netherlands Organization for Scientific Research (NWO).

References

- Alexander GE, DeLong MR, Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*. 1986; 9:357–381.
- Alexander GE, Crutcher MF, DeLong MR. Basal ganglia–thalamocortical circuits: Parallel substrates for motor, oculomotor, ‘prefrontal’ and ‘limbic’ functions. *Progress in Brain Research*. 1990; 85:119–147. [PubMed: 2094891]
- Arnsten AFT. Catecholamine modulation of prefrontal cortical cognitive function. *Trends in Cognitive Sciences*. 1998; 2(11):436–447. [PubMed: 21227275]
- Balleine BW, Delgado MR, Hikosaka O. The role of the dorsal striatum in reward and decision-making. *Journal of Neuroscience*. 2007; 27:8161–8165. [PubMed: 17670959]
- Bertsekas, DP.; Tsitsiklis, JN. *Neuro-dynamic programming*. Athena Scientific; Belmont, MA: 1996.
- Bjorklund A, Dunnett SB. Dopamine neuron systems in the brain: An update. *Trends in Neurosciences*. 2007; 30(5):194–202. [PubMed: 17408759]
- Bódi N, Kéri S, Nagi H, Moustafa A, Myers CE, Daw N, et al. Reward-learning and the novelty-seeking personality: A between- and within-subjects study of the effects of dopamine agonists on young Parkinson's patients. *Brain*. 2009; 132(9):2385–2395. [PubMed: 19416950]
- Brown P, Marsden CD. What do the basal ganglia do? *Lancet*. 1998; 351(9118):1801. [PubMed: 9635969]
- Cools R. Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease. *Neuroscience and Biobehavioral Reviews*. 2006; 30:1–23. [PubMed: 15935475]
- Cools R, Barker RA, Sahakian BJ, Robbins TW. Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. *Cerebral Cortex*. 2001; 11(12):1136–1143. [PubMed: 11709484]
- Cools R, Frank MJ, Gibbs SE, Miyakawa A, Jagust W, D'Esposito M. Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. *Journal of Neuroscience*. 2009; 29(5):1538–1543. [PubMed: 19193900]
- Cooper JA, Sagar HJ, Doherty SM, Jordan N, Ridswell P, Sullivan EV. Different effects of dopaminergic and anticholinergic therapies on cognitive and motor function in Parkinson's disease. *Brain*. 1992; 115(6):1701–1725. [PubMed: 1486457]
- Crevoisier C, Monreal A, Metzger B, Nilsen T. Comparative single- and multiple-dose pharmacokinetics of levodopa and 3-O-methyldopa following a new dual-release and a conventional slow-release formulation of levodopa and benserazide in healthy volunteers. *European Neurology*. 2003; 49(1):39–44. [PubMed: 12464717]
- Czernecki V, Pillon B, Houeto J, Pochon J, Levy R, Dubois B. Motivation, reward, and Parkinson's disease: Influence of dopatherapy. *Neuropsychologia*. 2002; 40(13):2257–2267. [PubMed: 12417456]
- Dayan P, Kakade S, Montague P. Learning and selective attention. *Nature Neuroscience*. 2000; 3:S1218–S1223.
- Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*. 2005; 8(12):1704–1711.
- Eyni Y, Horvitz J. Opposing roles of D1 and D2 receptors in appetitive conditioning. *Journal of Neuroscience*. 2003; 23(5):1584–1587. [PubMed: 12629161]
- Faure A, Haberland U, Condé F, El Massioui N. Lesion to the nigrostriatal dopamine system disrupts stimulus–response habit formation. *Journal of Neuroscience*. 2005; 25:2771–2780. [PubMed: 15772337]
- Folstein M, Folstein S, McHugh P. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12(3):189–198. [PubMed: 1202204]
- Frank MJ. Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience*. 2005; 17(1):51–72. [PubMed: 15701239]

- Frank MJ, Seeberger LC, O'Reilly RC. By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science*. 2004; 306:1940–1943. [PubMed: 15528409]
- Gasser UE, Jorga K, Crevoisier C, Hovens SEL, van Giersbergen PLM. COMT inhibition by tolcapone further improves levodopa pharmacokinetics when combined with a dual-release formulation of levodopa benserazide – A novel principle in the treatment of Parkinson's disease. *European Neurology*. 1999; 41(4):206–211. [PubMed: 10343151]
- Gerardin E, Lehericy S, Pochon JB, Tezenas du Montcel S, Mangin JF, Poupon F, et al. Foot, hand, face and eye representation in the human striatum. *Cerebral Cortex*. 2003; 13:162–169. [PubMed: 12507947]
- Goldman-Rakic PS, Muly EC II, Williams GV. D1 receptors in prefrontal cells and circuits. *Brain Research Review*. 2000; 31:295–301.
- Gotham A, Brown R, Marsden C. 'Frontal' cognitive function in patients with Parkinson's disease 'on' and 'off' levodopa. *Brain*. 1988; 111(2):299–321. [PubMed: 3378138]
- Graef SGS, Biele G, Krugel LK, Marzinzik F, Wahl M, Wotka J, et al. Differential influence of levodopa on reward-based learning in Parkinson's disease. *Frontiers in Human Neuroscience*. 2010; 4:169. [PubMed: 21048900]
- Haruno M, Kawato M. Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus–action reward association learning. *Journal of Neurophysiology*. 2006a; 95(2):948–959. [PubMed: 16192338]
- Haruno M, Kawato M. Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus–action–reward association learning. *Neural Networks*. 2006b; 19(8):1242–1254. [PubMed: 16987637]
- Hornykiewicz O. Mechanisms of action of L-DOPA in Parkinson's disease. *Life Sciences*. 1974; 15(7):1249–1259. [PubMed: 4620977]
- Kaasinen V, Rinne JO. Functional imaging studies of dopamine system and cognition in normal aging and Parkinson's disease. *Neuroscience and Biobehavioral Reviews*. 2002; 26(7):785–793. [PubMed: 12470690]
- Kimura M. Role of basal ganglia in behavioral learning. *Neuroscience Research*. 1995; 22:353–358. [PubMed: 7478299]
- Kish SJ, Shannak K, Hornykiewicz O. Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease: Pathophysiologic and clinical implications. *New England Journal of Medicine*. 1988; 318:876–880. [PubMed: 3352672]
- Knowlton BJ, Mangels JA, Squire LR. A neostriatal habit learning system in humans. *Science*. 1996; 273:1399–1402. [PubMed: 8703077]
- Knutson B, Adams C, Fong G, Hommer D. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience*. 2001; 21(16):1–5.
- Koepp MJ, Gunn RN, Lawrence AD, Cunningham VJ, Dagher A, Jones T, et al. Evidence for striatal dopamine release during a video game. *Nature*. 1998; 393:266–268. [PubMed: 9607763]
- Lezak, MD. *Neuropsychological assessment*. 3rd ed.. Oxford University Press; New York: 1995.
- Maddox W, Ashby F. Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioral Processes*. 2004; 66(3):309–332.
- McAuley JH. The physiological basis of clinical deficits in Parkinson's disease. *Progress in Neurobiology*. 2003; 69:27–48. [PubMed: 12637171]
- McClure S, Berns G, Montague P. Temporal prediction errors in a passive learning task activate human striatum. *Neuron*. 2003; 38(2):339–346. [PubMed: 12718866]
- O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*. 2004; 304(5669):452–454. [PubMed: 15087550]
- O'Reilly R, Frank M. Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*. 2006; 18(2):283–328. [PubMed: 16378516]
- Oyama K, Hernadi I, Iijima T, Tsutsui K-I. Reward prediction error coding in dorsal striatal neurons. *Journal of Neuroscience*. 2010; 30:11447–11457. [PubMed: 20739566]

- Packard M, Knowlton B. Learning and memory functions of the Basal Ganglia. *Annual Review of Neuroscience*. 2002; 25:563–593.
- Pappata S, Dehaene S, Poline JB, Gregoire MC, Jobert A, Delforge J, et al. In vivo detection of striatal dopamine release during reward: A PET study with [¹¹C] Raclopride and a single dynamic scan approach. *NeuroImage*. 2002; 16:1015–1027. [PubMed: 12202089]
- Rowe JB, Hughes L, Ghosh BCP, Eckstein D, Williams-Gray CH, Fallon S, et al. Parkinson's disease and dopaminergic therapy and dopaminergic therapy – Differential effects on movement, reward, and cognition. *Brain*. 2008; 131:2094–2105. [PubMed: 18577547]
- Rutledge RB, Lazzaro SC, et al. Dopaminergic drugs modulate learning rates and preservation in Parkinson's patients in a dynamic foraging task. *Journal of Neurosci*. 2009; 29(48):15104–15114.
- Schönberg T, Daw ND, Joel D, O'Doherty JP. Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *Journal of Neuroscience*. 2007; 27:12860–12867. [PubMed: 18032658]
- Schultz W. Getting formal with dopamine and reward. *Neuron*. 2002; 36(2):241–263. [PubMed: 12383780]
- Schultz W, Tremblay L, Hollerman JR. Changes in behavior-related neuronal activity in the striatum during learning. *Trends in Neurosciences*. 2003; 26(6):321–328. [PubMed: 12798602]
- Seger CA, Cincotta CM. The roles of the caudate nucleus in human classification learning. *Journal of Neuroscience*. 2005; 25(11):2941–2951. [PubMed: 15772354]
- Seymour B, Daw N, Dayan P, Singer T, Dolan R. Differential encoding of losses and gains in the human striatum. *Journal of Neuroscience*. 2007; 27(18):4826–4831. [PubMed: 17475790]
- Shohamy D, Myers CE, Grossman S, Sage J, Gluck MA. The role of dopamine in cognitive sequence learning: Evidence from Parkinson's disease. *Behavioural Brain Research*. 2005; 156(2):191–199. [PubMed: 15582105]
- Shohamy D, Myers CE, Kalanithi J, Gluck MA. Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience and Biobehavioral Reviews*. 2008; 32(2):219–236. [PubMed: 18061261]
- Sutton, RS.; Barto, AG. *Reinforcement learning: An introduction*. MIT Press; Cambridge, MA: 1998.
- Swanson R, Rogers RD, Sahakian BJ, Summers BA, Polkey CE, Robbins TW. Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: Possible adverse effects of dopaminergic medication. *Neuropsychologia*. 2000; 38(5):596–612. [PubMed: 10689037]
- Tricomi E, Delgado M, Fiez J. Modulation of caudate activity by action contingency. *Neuron*. 2004; 41(2):281–292. [PubMed: 14741108]
- Voon V, Pessiglione M, et al. Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. *Neuron*. 2010; 65(1):135–142. [PubMed: 20152119]
- Weintraub D, Siderowf AD, Potenza MN, Goveas J, Morales KH, Duda JE, et al. Association of dopamine agonist use with impulse control disorders in Parkinson disease. *Archives of Neurology*. 2006; 63(7):969–973. [PubMed: 16831966]
- Wylie SA, van den Wildenberg WPM, Ridderinkhof KR, Bashore TR, Powell VD, Manning CA, et al. The effect of Parkinson's disease on interference control during action selection. *Neuropsychologia*. 2009a; 47(1):145–157. [PubMed: 18761363]
- Wylie SA, van den Wildenberg WPM, Ridderinkhof KR, Bashore TR, Powell VD, Manning CA, et al. The effect of speed-accuracy strategy on response interference control in Parkinson's Disease. *Neuropsychologia*. 2009b; 47:1944–1853. [PubMed: 19428427]
- Yin HH, Knowlton BJ, Balleine BW. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience*. 2004; 19:181–189. [PubMed: 14750976]
- Young, P. *Recursive estimation and time series*. Springer-Verlag; New York: 1984.

Fig. 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 to this specific stimulus.

Fig. 2.

(a) Cumulative accuracy for the 90/10 Condition by Medication (On/Off). (b) Cumulative accuracy for the 80/20 Condition by Medication (On/Off). (c) Cumulative accuracy for the 70/30 Condition by Medication (On/Off).

Fig. 3.

(a) Mean RPE values from the first 24 trials separate for each condition. (b) Mean RPE values from the second 24 trials separate for each condition.

Fig. 4.
(a) Mean SADR values from the first 24 trials separate for each condition. (b) Mean SADR values from the second 24 trials separate for each condition.

Fig. 5. Δ RPE (ON–OFF) in the 90/10 Condition as a function of LEDD. Δ SADRP (ON–OFF) in the 90/10 Condition as a function of LEDD.

Table 1**A. PD patient information.**

	Sample $N = 20$	
	Mean	SE
Age (yrs)	68.5	± 1.6
Sex (male/female)	14/6	
MMSE	28.7	± 0.3
Years since disease onset	8.08	± 1.3
L-Dopa (daily dose mg)	527.5	± 54.3
Agonist dose in LEDD (daily dose mg)	106.8	± 39.0
Total LEDD (daily dose mg)	634.3	± 74.7
Finger tapping ON (# taps)	39.5	± 2.2
Finger tapping OFF (# taps)	40.3	± 2.0
Pegboard ON (time in s)	32.3	± 2.3
Pegboard OFF (time in s)	33.5	± 2.0

Table 2

Δ SADRP and Δ RPE (ON compared to OFF medication) and SADRPs OFF and RPE OFF medication are correlated with LEDD, disease duration and age.

Variables	LEDD	Disease duration	Age
LEDD (daily mg)	1	0.82 ^{**}	-0.01
Disease duration (yrs)	0.82 ^{**}	1	0.29
Age (yrs)	-0.01	0.29	1
Δ RPE 90/10 condition	0.48 [*]	0.20	0.16
Δ RPE 80/20 condition	-0.08	-0.08	0.01
Δ RPE 70/30 condition	0.31	0.20	-0.20
Δ SADRP 90/10 condition	-0.66 ^{**}	-0.43	0.09
Δ SADRP 80/20 condition	0.01	-0.02	-0.02
Δ SADRP 70/30 condition	-0.002	0.07	-0.02
RPE 90/10 OFF	-0.25	-0.31	-0.25
RPE 80/20 OFF	0.04	0.02	0.03
RPE 70/30 OFF	-0.28	-0.04	-0.29
SADRP 90/10 OFF	0.34	0.19	-0.29
SADRP 80/20 OFF	-0.17	-0.15	-0.21
SADRP 70/30 OFF	0.20	-0.16	-0.48 [*]

*
 $p < 0.05$.

**
 $p < 0.01$.