Intact action segmentation in Parkinson's disease: hypothesis testing using a novel computational approach

Anne-Marike Schiffer[1]*, Alejo J. Nevado-Holgado[1,2], Andreas Johnen[3], Anna R. Schönberger[4], Gereon R. Fink[4,5] and Ricarda I. Schubotz[4,6]

[1]Department of Experimental Psychology, University of Oxford, Oxford, UK

[2]Department of Psychiatry, University of Oxford, Oxford, UK

[3]Department of Neurology, University Hospital Münster, Münster, Germany

[4]Department of Neurology, University Hospital Cologne, Cologne, Germany

[5]Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM3), Research Centre Jülich, Jülich, Germany

[6]Biological Psychology, Department of Psychology, Westfälische-Wilhelms Universität Münster, Münster, Germany

*Corresponding author: anne-marike.schiffer@psy.ox.ac.uk

Abstract

Action observation is known to trigger predictions of the ongoing course of action and thus considered a hallmark example for predictive perception. A related task, which explicitly taps into the ability to predict actions based on their internal representations, is action segmentation; the task requires participants to demarcate where one action step is completed and another one begins. It thus benefits from a temporally precise prediction of the current action. Formation and exploitation of these temporal predictions of external events is now closely associated with a network including the basal ganglia and prefrontal cortex.

Because decline of dopaminergic innervation leads to impaired function of the basal ganglia and prefrontal cortex in Parkinson's disease (PD), we hypothesised that PD patients would show increased temporal variability in the action segmentation task, especially under medication withdrawal (hypothesis 1).

Another crucial aspect of action segmentation is its reliance on a semantic representation of actions. There is no evidence to suggest that action representations are substantially altered, or cannot be accessed, in non-demented PD patients. We therefore expected action segmentation judgments to follow the same overall patterns in PD patients and healthy controls (hypothesis 2), resulting in comparable segmentation profiles. Both hypotheses were tested with a novel classification approach.

We present evidence for both hypotheses in the present study: classifier performance was slightly decreased when it was tested for its ability to predict the identity of movies segmented by PD patients, and a measure of normativity of response behaviour was decreased when patients segmented movies under medication-withdrawal without access to an episodic memory of the sequence. This pattern of results is consistent with hypothesis 1. However, the classifier analysis also revealed that responses given by patients and controls create very similar action-specific patterns, thus delivering evidence in favour hypothesis 2.

In terms of methodology, the use of classifiers in the present study allowed us to establish similarity of behaviour across groups (hypothesis 2). The approach opens up a new avenue that standard statistical methods often fail to provide and is discussed in terms of its merits to measure hypothesised similarities across study populations.

Acknowledgements:
The authors thank Kim Hannah Krause and Robert Schnuerch for their help in data collection.

Conflict of Interest:
The authors declare no competing financial interests.

Funding:

This study was supported by the German Research Foundation (DFG), Clinical Research Group (KFO)219, TP4 (GZ SCHU 1439/3-1).

1 Introduction

Parkinson's disease (PD) is a condition with well-defined neurological changes. It results from a loss of dopaminergic cells in the substantia nigra (Bernheimer et al., 1973; Birkmayer and Wuketich, 1976), which leads to decreased levels of this neurotransmitter in the basal ganglia and the prefrontal cortex (PFC). PD is signified by prominent motor impairments such as tremor, bradykinesia, and rigor. These motor symptoms are often accompanied by cognitive changes, including compromised ability to learn from feedback and limited use of the predictability of external events (Flowers, 1978; Cameron et al., 2010; Cools et al., 2003; Cools et al., 2001; Cools, 2006; Crawford et al., 1989; Frank, 2006; Zalla et al., 1998; Shohamy et al., 2008*)*. A related impairment in PD which has recently been linked to the basal ganglia and the prefrontal cortex is the internally driven prediction of external events (Schönberger, et al., 2013).

1.1 (Temporal) Prediction in a basal ganglia network

The proposal that the basal ganglia are involved in prediction of the content and temporal onset of external events (referred to as sensory states in the original literature Bischoff-Grethe, Crowley, and Arbib, 2003) is grounded in a combination of findings from patient data with data from animal, imaging, and modelling research (Alm, 2004; Balleine, Liljeholm, and Ostlund, 2009; Berns and Sejnowski, 1998; Bischoff-Grethe, Crowley, and Arbib, 2003; Schönberger, et al., 2013). The research suggests that the basal ganglia and prefrontal cortex, and particularly the supplementary motor area (SMA), work in concert in learning, selecting, and timing predictions of external events (Lewis et al., 2004; Stocco, Lebiere, & Anderson, 2010; Schiffer, Wasak, & Yeung, 2015; Schönberger, et al., 2013; see Coull & Nobre, 2008 for a dissenting view). Because decline of dopaminergic innervation of the basal ganglia and prefrontal cortex is a hallmark feature of PD, this research suggests that PD patients should be compromised in the fast prediction of event sequences, particularly under medication withdrawal. The present study tested this hypothesis explicitly, implementing an action segmentation task.

1.2 Action segmentation requires exploitation of semantic knowledge and benefits from prediction of forthcoming events

In the segmentation task participants observe an actor performing familiar activities and are required to indicate their subjective judgment whether an action boundary has occurred, i.e.,

whether an action step has been completed and a new action step has been initiated. These segmentation judgments, also referred to as boundary detection reports, are usually given in the form of a button press (Zacks et al., 2001; Schubotz et al., 2012; Baldwin et al., 2008; Newtson and Engquist, 1976). Because actions are highly structured and action observation is known to trigger online predictions of forthcoming action steps (Csibra, 2007; Colder, 2011; Botvinick and Plaut, 2004; Kilner, Friston, and Frith, 2007; Kilner et al., 2004; Schiffer et al., 2013; Stadler et al., 2011), reliable and fast performance in action-segmentation tasks requires two core abilities:

First, action segmentation benefits from the ability to generate a temporally precise prediction of the course of the current action, including the end of one action step and the beginning of the next action step thereafter. Detection of stimuli is not only aided by predictability of occurrence, but also additionally facilitated by predictability of stimulus onset (Rohenkohl et al., 2012). Thus, predicting which action step is to follow, and at what time this action step would naturally commence, aids boundary detection in the action segmentation task.

Importantly, if the basal ganglia are involved in real-time prediction of sequential events (Schiffer & Schubotz, 2011), we would expect increased variability in the timing of the response around action boundaries (Baldwin et al., 2008; Newtson and Engquist, 1976) in PD patients. The action-segmentation paradigm thus provides a sensitive test for the hypothesis that compromised dopaminergic innervation of the basal ganglia and prefrontal cortex leads to increased temporal variability in response behaviour, particularly under medication withdrawal (hypothesis 1), indicating impaired (temporal) prediction and delayed assessment of forthcoming sensory states.

A second, profound aspect of action segmentation is that observers have to rely on an internal representation of the single steps that together form specific actions (action semantics) to detect the end of one action step and the beginning of another. Some authors have argued that PD patients should be impaired in action segmentation (Zacks & Sargent, 2010). However, while learning and retrieval of action semantics has repeatedly been shown to involve a fronto-parietal network extending to the temporal lobes (Decety et al., 1997; Spunt, Falk, and Lieberman, 2010; Watson and Chatterjee, 2011; Hoffman, Jones, and Ralph, 2012; Schubotz et al., 2012; Schiffer et al., 2013), evidence for an involvement of the basal ganglia is missing. We therefore propose that the ability to segment actions should be largely intact in non-demented PD (hypothesis 2), resulting in comparable segmentation profiles.

1.3 Assessing action segmentation components in a patient study

We tested these hypotheses in a cohort of patients with idiopathic Parkinson's disease and a group of age-matched controls. To assess whether changes in dopamine availability exert an effect on the ability to segment actions per se and increase the temporal variability of segmentation behaviour, PD patients underwent two experimental sessions, one with their usual dopamine replacement therapy unchanged (ON) and one under withdrawal of their dopamine replacement therapy (OFF). Healthy controls took part in two separate sessions without medication. Their **virtual** medication status (pseudo ON and OFF status) was yoked to the random order of ON and OFF tests in the matched PD patients. During each session, participants segmented a different set of 6 multi-step action movies twice, allowing comparison of segmentation reliability under different medication status.

1.4 Classification approach to assess similarity

Predictions of similarity, central to our second hypothesis, are statistically challenging, because inference statistic measures aim at establishing differences between groups. Even if these measures fail to establish a difference between groups or conditions, such null effects cannot be taken as a proof of similarity (Cohen, 1994). Moreover, our hypotheses demand an estimate of the exact *degree* of similarity between response patterns. We resolved this paradox by developing a novel methodology, which implements a computational classifier. To show that PD patients and healthy controls can rely on the same action models, we transformed their response behaviour in the action-segmentation task into a temporal profile of response probability, expressed as the function that represents the probability to make a response for each moment in time. Bringing the data into this format allowed us to use these temporal response profiles in a computational classifier (**Figure 1;** please refer to Methods section 2.2 and 2.4.1 for further explanation).

We trained a classifier to predict movie identity using the data from a subset of participants as a training set and another subset of participants as a test set. The hypothesised above-chance classification of movie-specific response profiles when testing data and training data are taken from different groups strongly indicates behavioural similarity. This behavioural similarity is evidence in favour of intact semantic representation of action structure in PD. At the same time, the predicted differences in classification performance between different (above-chance) cross-group classifications would show the predicted differences in the temporal precision of segmentation behaviour in PD.

2 Materials and Methods

2.1 Participants

A total of 32 male participants took part in the experiments: 16 patients with idiopathic Parkinson's Disease (PD) and 16 healthy controls, individually matched for age, handedness, and education (please refer to **Table 1** for a summary of the information on the patient and control population). Every invited PD patient was tested and no dataset was discarded. For patients to be invited and included in the study they had to fulfil the following list of inclusion criteria. Patients had to be diagnosed with idiopathic Parkinson's disease. They had to be aged between 18 and 80, have given written informed consent, and weren't allowed to take part in any other study on the same day. Lastly, testing in their medication OFF state was conducted within their regular, scheduled assessment, during which they withdraw from their individual medication to test for symptom severity and dopa-responsiveness. Thus, the patients were not in their medication OFF as part of a clinical trial.

Exclusion criteria were: receiving deep-brain stimulation and suffering from further neurological or life-expectancy limiting diseases. Inclusion/exclusion criteria for matched controls were comparable, except for the presence of idiopathic Parkinson's disease, or any other neurological or psychiatric condition, which were exclusion criteria for control participants. There was also no relationship to a scheduled stay at the hospital for the control group, as these participants did not receive or withdraw from any medication.

All participants had an introductory session one day before the first test session to practise a short version of the main task and control tasks. This practice session did not contain any of the videos that were later used in the real test sessions (ON or OFF). The purpose of this pilot session was to ensure that all participants would understand the tasks, even if their first test session took place under medication withdrawal. One matched control had to be replaced by another equally well-matched control participant, as the first person did not understand the instructions of various subtasks.

Average Unified Parkinson's Disease Rating Scale (UPDRS) scores for healthy controls was 1.15, compared to 23.8 for PD patients (mean ON medication: 20.9, mean OFF medication: 27.1). The difference in UPDRS scores between PD patients and controls was highly significant in a one-sided t-tests ($T = 17.8$, $p < 10^{-18}$ $df = 31$), and so was the difference between ON and OFF session for PD patients ($T = 6.1$, $p < 10^{-6}$, $df = 15$). The

average Parkinson Neuropsychometric Dementia Assessment (PANDA) scores were 25.4 and 26 for PD patients and healthy controls, respectively. Beck's Depression Inventory (BDI) scores were 9.7 vs. 5.75 for PD patients and healthy controls, respectively. The differences in PANDA and BDI scores were not statistically significant in one-tailed t-tests (PANDA: T= 0.5, p = 0.31; BDI: T = 0.8, p = 0.21). No healthy control participant and no PD patient scored lower than 14 points, indicating that that no participant fulfilled the cut-off for dementia. All but one participant scored higher than 18 points, indicating age-appropriate function (Kalbe et al., 2008). One PD patient scored 16 points, thus being in the range of subtle cognitive impairment. The proceedings were approved by the local ethics committee of the Medical Faculty of the University of Cologne and the work described was carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving human subjects.

2.2 Task

2.2.1 Action segmentation task:

All participants took part in two experimental sessions. For the PD patients, one session took place when they were on their individual, regular dopamine-replacement medication (ON session), and another session after over-nightly medication withdrawal (OFF session). The order of ON/OFF sessions was randomised across patients. An overview of medication specifics is included in Table 4. For the healthy controls, whether a session was assigned ON or OFF status was yoked to their matched patient's order of sessions. Note that healthy controls did not receive any dopaminergic medication in any session. Therefore, these sessions will henceforth be described as pseudo ON/pseudo OFF, to emphasize that no medication was involved at any stage for the healthy volunteers.

Within each test session, the participant segmented 6 different short movies of naturalistic action sequences 2 times each (please refer to Table 3 for a description of the movies)**.** The first and second segmentation instance within sessions included the same movies, but no movie was repeated in the next session. The selection of the 6 movies for each of the first session was pseudo-randomised and the second session contained the other 6 movies of the set of 12. Pseudo-randomisation ensured that each of the 12 movies appeared in all possible conditions across participants: first sessions ON medication, first sessions OFF medication, second sessions ON medication and second sessions OFF medication. This setup allowed us to measure reliability scores and movie specific segmentation independent of order and

medication effects (please refer to Schubotz et al., 2012 for a comparable design in a study with young healthy volunteers).

Within the segmentation task, participants were instructed to indicate with a button press whenever a new action step began. In more detail, participants were told to press a button when they felt (emphasis on the subjectivity of the judgement) that one action step had finished and a new action step was to begin (example judgments for the two segmentations performed on the same movie., ie, within one session, are depicted by the blue lines and bars in Figure 1). They were told that an action step might relate to what they would say if we asked them to give an online record of the actions they saw to a bystander**.** Responses were made with a standard QWERTZ keyboard, by pressing the space bar.

2.2.2 Motor control task:

Participants' motor behaviour was assessed in a separate task. In this part of the experiment, subjects were presented with a stream of white and red squares on a grey-background monitor at 1/5 Hz. Their task was to respond as quickly as possible to the red crosses, while ignoring the white ones. Each colour appeared equally often in a randomised order. The task was run for 60 trials, i.e., 30 target trials (red crosses). Crosses were presented in font size 30. Responses were made with a standard QWERTZ keyboard, by pressing the space bar.

2.2.3 Cognitive control tasks:

To increase the interpretability of the classifier results we conducted a number of control tasks which tested for differences between patients and healthy controls in: the ability to retrieve semantically associated items, the ability to recognize a familiar action episode, and in the ability to predict the on-going course of an action.

Semantic association control task: The ability to retrieve semantically associated items was tested in a paradigm in which participants were presented with a pair of nouns, e.g., "sugar","flour", and had to name a related item, e.g., "salt". Reaction times were recorded over 10 trials per session, with a microphone that was sensitive to speech onset. Participants had up to 6 seconds to initiate their response. The inter-trial interval was 1 second. Correctness of the 10 responses (i.e., whether the participants response was semantically related to the word pair) was later rated by two independent observers. These were blind to disease status and medication.

Episodic recognition control task: The ability to recognize a familiar action was tested on another set of 10 everyday action movies (not appearing in the segmentation tasks),

which were presented at the beginning and end of each experimental session. These movies contained short everyday actions, all performed while sitting at a table, such as preparing muesli, stapling a stack of paper, wrapping up a parcel, etc. (please refer to Schiffer et al., 2013 for pictures showing some of the actions). When participants saw the movies again at the end of the test session, movies either appeared in the same version as before or in a different version (please refer to Schiffer et al., 2012; 2013 for more details). Participants had to press one of two response buttons (left arrow key and down arrow key on a standard QWERTZ keyboard) to indicate whether the movie had been presented as before. Participants had up to 6 seconds to initiate their response. The inter-trial interval was 1 second.

Action prediction/association control task: Lastly, to test for participants' ability to predict a likely on-going course of action, participants were presented with a third set of 10 movies, which ended abruptly after the completion of an action step. These movies were again taken from the sample implemented in Schiffer et al. (2012; 2013), showing everyday actions taking place at a table; there was no overlap between the movies used for any of the control tasks within subjects. Participants were then instructed to name a probable next action step. Voice responses were again recorded with a microphone that was sensitive to the time point of speech onset. Participants had up to 6 seconds to initiate their response. The inter-trial interval was 1 second. Please note that while prediction of likely next action steps would help to decrease reaction times in this task, timing of these associated predictions is not as crucial as in the action segmentation paradigm.

2.3 Descriptive statistics

In a first simple analysis, we used number of segmentations as an approximate measure to estimate the reliability of segmentation responses. The number of segmentations for each movie was correlated within each session for each participant to yield average correlation scores across all six respective movies for each participant in each session (cmp. Schubotz et al., 2012).

2.3.1 Segmentation agreement

In a next step, we assessed how normative segmentation judgments were (i.e., how much a the segmentation profile of a movie derived from one person was in agreement with how other participants segment the same movie). This variable needs to be as closely related to the timing of segmentation judgments as possible, as this approach complements the classifier

analysis (section 2.4). To obtain a normativity score, we first established a symmetric time window around each segmentation judgment 'a'. We then counted how many other times segmentation judgments were placed within this window around 'a' by the other participants.

To avoid any bias, we excluded the judgments by the participant in their second segmentation instance of the same movie and the judgments by his matched control. We call this the number of *segmentation agreements* for segmentation judgment 'a'. This represents a statistical random variable which measures how normative a given segmentation judgment 'a' is. Therefore, we can use this random variable to estimate how much the segmentations produced by a given group (e.g., PD patients OFF medication) agree with the general population. A group including participants who segment a movie in a manner different from the average population will get lower agreement scores. Conversely, a group with participants that segment more normatively will get higher agreement scores (see Kurby, Asiala, and Mills, 2014 for a closely related approach).

In addition to the inference-statistic measures and the normativity estimate, we also employed a classifier approach to test whether PD patients rely on the same semantic structure (i.e., are uncompromised in their ability) to segment actions. The classifier approach extends the possibilities of classic inference statistics; while classic approaches test for the difference between populations, classifiers can show that the data drawn from one sample can predict the shape of the data of the corresponding sample - a strong argument in favour of similarity.

2.4 Within-and between-groups classification

The power of a classifier analysis is its ability to predict the category of an item based on information the classifier previously gathered about other items from all existing categories. Harnessing this characteristic, we devised a classifier analysis to show that classification in PD patients and healthy controls is so consistent that a classifier could predict which movie's data it was currently being presented with.

2.4.1 Preprocessing

Given a number of samples, of which each belongs to one of two possible classes, a classifier attempts to learn the underlining *sample-class mapping* (Murphy, 2012). Samples are N-dimensional vectors, while classes are labels with two possible values {class 1, class 2}. In the present study, the task of the classifier was to assign the identity ("correct name") to each

movie, based on the segmentation judgments. This means that the segmentation judgments served as N-dimensional *samples*, and the *classes* were the correct name of the movie. However, the segmentations do not have a constant number of dimensions, as each participant may make a different number of segmentation judgments in the same movie (i.e., participants responded more or less often for each movie). To achieve the same vector length for each sample (i.e., each segmentation instance for each movie for every participant), we used a Fourier approach which, given a movie and a subject, obtained the probability of the subject placing a segmentation judgment at any time point for that movie, in essence a temporal profile of the typical response behaviour (this smooth probability function for the example movie is depicted in the red line in **Figure 1**). This probability function has a fixed number of dimensions (each time point is a dimension). In more detail: using formal nomenclature, the segmentation response of subject S, when watching movie M in trial T is *eSMT* (*t*), and can be described as a sequence of δ-dirac functions (δ functions are also commonly referred to as stick functions): $e_{SMT} = \begin{pmatrix} 1 & segmentation\ at\ time\ t \\ 0 & otherwise \end{pmatrix}$

A smooth probability density function (i.e., pSM(t)) is the natural result of representing a function of time with only the first few components of its Fourier transform (Diniz et al., 2010). This function estimates the probability of the subject pressing the segmentation button at time t for that given movie. The following four steps were implemented to derive this function: In a first step, we calculated the Fourier transform of *eSMT* (*t*):

$$E_{SMT}(f) = \sum_{t=1} \exp(-2\pi f t)$$

where f are the different Fourier components, evaluated at frequencies 1Δf, 2Δf...,with Δf = 1000 divided by the total duration of the movie. In simple terms, Fourier transforms allow to generate a soft approximation of the signal described in the sets of δ functions. In the next step, we picked only the first 8 components of this transform to achieve a smooth representation. We chose 8 components because this provided time profiles that were smooth enough for the averages to converge. However, getting a few more or less components did

not change the results of the overall analysis. Only using either very few components (<4) or too many (>20), will hampered the classifier's performance - and it is then impaired in all conditions (for PD and controls), as the time profiles will change either too slowly with time (for < 4 all movies will render the same time profile) or too fast (for > 20 different segmentation profiles of the same movie will start to diverge). Third, for each subject and movie, we averaged these 8 components across trials:

$$P_{SM}(f) = \sum_{T} E_{SMT}(f)$$

The fourth and last step was to apply the inverse Fourier transform to obtain the temporal profile of this signal:

$$P_{SM}(t) = \sum_{f=1\Delta f, 2\Delta f, \ldots}^{8\Delta f} \exp(2\pi f t)$$

As we eliminated the elements containing the high frequency components of the original δ functions, we obtained a smooth version of the segmentation times (this is a general property of the Fourier transform and of low-pass filters). Assuming that the probability of pressing the segmentation button changes slowly over time, this effectively created an estimation of this probability based on the *eSMT* samples (please refer to **Figure 1** for the depiction of a smooth probability-density function achieved in this way).

2.4.2 Classification

For each movie M, we selected 30 equidistant time points, with a separation equal to $^{1}/_{30}$ of the total length of that movie as input dimensions for the classifier. The purpose of the classifier was then to test whether it could assign the movie class (identity) correctly based on the information from these 30 dimensions (**Figure 3**). In simple words, the question is whether the classifier can, for example, identify that it is presented with the temporal profile of segmentations (segmentation pattern) of the movie that shows an actor doing the dishes

based on its training with the temporal profile of button-press probabilities for all movies, including the dishes movie.

This setup of movie-based classification allowed us to use the classifiers to measure how consistent participants within each group segmented movies. To this end, we iteratively selected one subject from the group and two movies, which served as the two classes that the classifier had to identify. We trained the classifier on all subjects (excluding the selected one), and measured whether it could correctly classify the probability-density function (temporal profile of response-probability) of the selected participant as one of the two movies. We repeated this leave-one-out training/testing procedure (also referred to as jack-knife approach) for all possible pairs of movies and for all participants in the given set of subjects. The obtained average number of correct classifications indicates how consistent the segmentation of movies was within this group of subjects.

A modification of this classification procedure allowed us to test how consistent segmentation is across two groups, A and B. To this end, we selected all the subjects of group A except for one as the traininig sample in the classifier, and tested the classifier's ability to predict movie identities for the matched subject of group B. This means, for example, that we trained the classifier with the segmentations from PD patients 2-16 and tested its ability to assign the correct label to segmentation patterns derived from the matched control of PD patient 1. The latter approach was used to measure whether the segmentations performed by PD patients (group A) were consistent with controls (group B).

3 Results

3.1 Descriptive statistics

In the segmentation task, PD patients segmented each action movie on average 10.4 times in their medication ON status and 9.8 times in their OFF status. Healthy controls segmented the same movies on average 12.2 times in the pseudo ON and 12.9 times in the pseudo OFF status. The time interval between two segmentation judgments was on average 9.5 seconds in ON status and every 10.5 seconds in OFF status. For the healthy controls, segmentation interval was on average 9.6 seconds in pseudo ON and 11.4 seconds in pseudo OFF. We analysed the number of segmentations for each group (PD/CONTROL) in each medication status (ON/OFF) using a repeated-measures ANOVA with between-subject factor GROUP and within-subject factor MEDICATION STATUS and found no significant main effect or

interaction (all $F_{(1,30)}<1$). These results indicate no strong differences in segmentation behaviour, i.e., PD patients did not segment significantly less often than controls, irrespective of medication status.

A correlation analysis was conducted on the number of responses for each movie and for each of the two instances of the segmentation task in each session, per participant. This yielded an average within-session segmentation-judgment reliability of r=.86 (p = 0.045) for PD patients ON medication, r=.87 (p = 0.039) OFF medication, r=.74 (p = 0.19) for healthy controls in pseudo ON, and r=.88 (p = 0.031) for healthy controls in pseudo OFF. We conducted a repeated-measures ANOVA on within-session correlation with the between-subject factor GROUP and within-subject factor MEDICATION STATUS and found no significant main effect or interaction (all $F(1,30)<1$). All correlation coefficients were Fisher z-transformed for group statistics.

3.1.1 Cognitive control tasks:

We analysed participants' reaction times and accuracy - measured as percent of correct responses - in 6 different repeated-measures ANOVAS (**Figure 4**). Each ANOVA contained the data from the patient population and their matched control (between-subject factor GROUP) under both medication conditions (within-subject factor MEDICATION STATUS). In the **Semantic association control task**, we found no significant main effect (all $F_{(1,30)}< 1$) of GROUP or MEDICATION STATUS and no significant interaction for accuracy rates. Reaction-time data likewise yielded no significant main effect (all $F_{(1,30)}< 1$) and no significant interaction.

We found no indication of a difference in accuracy in the **Episodic recognition control task**, with no significant main effects (all $F_{(1,30)}< 1$) and only a trend-level interaction of GROUP and MEDICATION STATUS ($F_{(1,30)} = 3.199$, p = 0.08). In the reaction-time data, we also found no significant main effect (GROUP $F_{(1,30)} = 1.3$, p = 0.26, MEDICATION STATUS $F_{(1,30)}< 1$) . There was no significant interaction ($F_{(1,30)} < 1$).

Finally, the **Action prediction/association control task** yielded a marginally significant effect of GROUP in the accuracy data ($F_{(1,30)} =3.84$, p = 0.059), but no main effect of MEDICATION STATUS and no interaction (both $F_{(1,30)} < 1$). In the reaction time data, we found no main effect (all $F_{(1,30)}< 1$) and no significant interaction ($F_{(1,30)} = 2.73$, p = 0.1). In sum, the results from the control tasks did not show a specific impairment in any group under any condition for functions which have to be considered necessary abilities for the action-segmentation task: the ability to retrieve associations in general and in relation to actions, and

the ability to learn about new action episodes. The latter may be necessary to engage in a compensatory strategy, as we will discuss later on.

The number of trials in all control tasks was very limited to reduce the time spent under medication withdrawal. This means that the test may have had not enough power to detect an impairment of function on the single-subject level. However, taken together with the results of the PANDA tests, which showed that no participant suffered from dementia (including associative learning and working memory abilities), and given that all participants performed extremely well (mean accuracy higher than 80% in all tasks), there is no compelling reason to assume that PD patients were impaired in action recognition, semantic retrieval, or episodic memory. These results permit no inferences on whether action recognition, semantic retrieval, or episodic memory *can* be impaired in PD. But they suggest that in the present population differences in behaviour established in the analysis of segmentation agreement and the classifier analysis were not driven by substantial impairments in these functions.

3.1.2 Segmentation agreement

The above reported analyses of segmentation frequency per movie and within-session correlation coefficients for segmentation frequency show that PD patients display consistency in their segmentation behaviour across ON and OFF status. At the same time, it is evident that the number of segmentations does not convey any information about segmentation location. In contrast, the following analysis and the classifier approach both used measures that were sensitive to the exact time-point of segmentation responses.

We used a time-window approach to measure within-group segmentation agreement. Given a segmentation judgment ‘a’, this approach measures how often other subjects also placed a segmentation judgment within a given time window around 'a'. This delivers a measure of normativity: when, for a given movie, a participant segments close to the time when many other subjects also make a segmentation judgment, the participant’s segmentation is in agreement with the population (see Methods and Figure 2 for details, and Kurby, Asiala, and Mills, 2014 for a closely related approach).

The histograms in Figure 5 show the segmentation agreement for PD patients and healthy controls in ON and OFF sessions, divided for the first and second segmentation instance for each movie. Interestingly, when PD patients were tested in their first session OFF medication, they showed significantly less agreement than control participants who

segmented a movie for the first time (Kolmogorov-Smirnov; p-value = 0.0061; ks-stat 0.073). In Figure 5 (lower left), this is evident because many segmentation judgments made by PD patients OFF medication in their first segmentation instance agree only with 10-30 segmentation judgments placed by other participants (i.e., only 10-30 other subjects placed a segmentation within the time window). However, there was no difference between groups' segmentation agreement the second time they segmented the movie. Tested ON medication, PD patients did not differ from healthy controls in their segmentation agreement scores for both the first and second segmentation (regardless the width of the time-window). The results shown in Figure 5 are based on a time window of 1.5 seconds half-width. This result holds for all window widths between 1 and 2.3 seconds. No statistically significant performance decrement for PD patients in any medication or segmentation-instance condition with wider windows was observed.

3.2 Classifier analysis

We used a classifier analysis to assess how consistent segmentation patterns were within and across our four groups (PD patients ON vs. OFF medication, healthy controls in pseudo ON vs. OFF session). These classifications produced 16 averages as shown in **Figure 6**. Averages were calculated across all the possible leave-one-out splits of the data for the training-group-A/testing-group-B classification. All of these classification performances were higher than 80% and t-tests showed that all of them were significantly different from chance at $p < 10^{-14}$ (**Table 3**). This allows the first inference that the commonalities in segmentation patterns far outweighed the differences, as the classifier would otherwise have performed at chance level (it would have "guessed" movie identity).

To test for any possible effect of training group, testing group, or medication status, we ran a 4-way ANOVA with the factors: (i) TRAINING GROUP (PD/CONTROL), (ii) TESTING GROUP (PD/CONTROL), (iii) MEDICATION STATUS TRAINING GROUP (ON/OFF), and (iv) MEDICATION STATUS TESTING GROUP (ON/OFF).

The first classifier did not include measures of motor impairment and classified solely on the dimensions derived from the smooth probability-density function for segmentation behaviour. This analysis yielded a significant main effect of TRAINING GROUP ($F_{(1,15)} = 6.99$; $p = 0.009$, a marginally significant main effect of TESTING GROUP ($F_{(1,15)} = 3.4$, $p = 0.066$, but no further main effect and no significant interaction. In a second classifier, we included standard deviation in reaction time in the motor control tasks as an additional dimension, to account for higher motor variability under dopamine-replacement withdrawal.

This step is necessary to link potential between-group differences to cognitive changes. This classifier showed again a main effect of TESTING GROUP ($F_{(1,15)}$= 12.39, p = 0.001), but no other main effect (all F<1, except main effect of training group at $F_{(1,15)}$= 1.15, p = 0.28), and no significant interaction (all F<1, except interaction of testing group by training group at $F_{(2,14)}$ = 1.44, p = 0.23).

Lastly, we repeated this second approach, using the standard deviation sigma from an ex-gaussian fit to the reaction-time data from the motor **control** task. Sigma in an ex-gaussian model of reaction-time data captures the amount of variance in the data. This analysis (**Figure 6**) likewise yielded a significant main effect of TESTING GROUP, ($F_{(1,15)}$ = 7.84, p = 0.001), but no other significant main effect or interaction.

4 Discussion

The present study investigated whether PD patients would display behavioural impairments in an action segmentation task, which requires the exploitation of structured semantic action representations and the generation and evaluation of predictions of forthcoming events. We expected that PD patients would show some temporal variability around segmentation points (1), but that the temporal pattern emerging from these segmentation points would be nearly indistinguishable between PD patients and healthy controls (2). We found evidence for both hypotheses in the present study. When participants were asked to segment action movies at meaningful boundaries, classifiers trained on the temporal pattern of segmentation responses were able to classify movie identity far above chance, for both training (PD or healthy controls) and testing groups (PD or healthy controls), under either medication status (ON/OFF). This core finding strongly suggests that PD patients have access to and exploit the same action knowledge as healthy controls in action segmentation.

As predicted by our first hypothesis (temporal variability), classifier performance was slightly decreased (while still far above chance) when it was tested for its ability to predict the identity of movies segmented by PD patients. This subtle change in performance indicated that PD patients' data contained more variability at segmentation points, thereby becoming marginally less predictable in classification. Importantly, this finding stands when motor variability, assessed in a separate motor control task, is accounted for by the classifier. Thus, this finding suggests that the difference between the two groups is caused by cognitive changes rather than a consequence of altered motor behaviour in PD. Notably and against expectations, this small deviation was not limited to a specific medication session.

Indeed, we found that segmentation in PD patients reached lower agreement scores only during the first of two segmentation instances in the OFF state. This lack of agreement with the average segmentation, or non-normativity, was not, however, present during the second segmentation instance in the OFF state, or any segmentation instance in the ON state. This striking pattern of a one-time-exposure training effect supports the idea that patients can use episodic memory for the content of the action sequence to compensate. Because we find this compensation in dopaminergic OFF state, it is likely to rely on a brain network that does not critically depend on dopaminergic innervation.

4.1 PD patients exploit the same action knowledge as healthy controls when segmenting action movies

Action segmentation relies on semantic action knowledge (Zacks et al., 2006; Kurby, Asiala, and Mills, 2014; Bailey et al., 2013). Learning and retrieving this action knowledge is associated with a network including the lateral prefrontal cortex and temporo-parietal areas (Binder et al., 2009; Buxbaum et al., 2007; Buxbaum, Kyle, and Menon, 2005). Recently, there has also been evidence for a hippocampal involvement (Schubotz et al., 2012), a region classically associated with episodic memory.

The putative role of the hippocampus is of particular interest since it is well established that although PD patients have difficulties to learn from (positive) feedback and compensate strategically for this impairment via explicit learning of stimulus-outcome contingencies (Shohamy et al., 2008). Learning response-outcome contingencies from feedback integration is assumed to rely on the basal ganglia and to involve the dopaminergic midbrain, while the suggested compensatory strategies are mediated by the hippocampus (Dagher et al., 2001; Shohamy et al., 2008;).

Clearly, attributing all compensatory function in PD to a hippocampal network is not warranted. This is not least because the hippocampus receives dense dopaminergic projection and the degree to which a potential decrease in innervation in PD could alter hippocampal function remains unclear. (Jay, 2003 for review) Further, it has been shown that hippocampal volume can be decreased in PD, especially in elderly patients and patients suffering from dementia (Brück et al., 2004; Camicioli et al., 2004; Churchyard & Lees, 1997 - please note that the PD patients in the present study did not suffer from dementia or memory problems). These findings suggest that hippocampal function may be impaired in PD, which could potentially have implications for the availability of hippocampal compensation mechanisms.

In contrast, the possibility that a hippocampal learning and memory mechanism may indeed be involved in compensation in this specific task is suggested by the episodic nature of the decrease in non-normativity: normativity scores in patients in the OFF status made a full recovery as soon as they had segmented the same movie one single time before**.** Lastly, the proposal that episodic memory can aid action segmentation and that this process is associated with the hippocampus receives some support from a study which showed non-normative segmentation behaviour in participants with decreased medial temporal lobe volume (Bailey et al., 2013). Thus, whether decrease in non-normative behaviour is in fact hippocampally mediated remains an open and exciting research question. An empirical study using classifiers to achieve a double dissociation between PD patients and patient groups with dementia would be highly desirable.

In light of the present results and our previous fMRI data (Schubotz et al., 2012), we propose that action segmentation based on action semantics and episodic memory relies on a network including prefrontal cortex (Grafman, 2003; Schubotz et al., 2012), cortical areas involved in action representation (Decety et al., 1997; Spunt, Falk, and Lieberman, 2010; Watson and Chatterjee, 2011; Hoffman, Jones, and Ralph, 2012), and the hippocampal formation (Schubotz et al., 2012 cf. Bailey et al., 2013). Intact dopaminergic innervation of the basal ganglia (and prefrontal cortex) does not appear essential for action segmentation, but is important for the precise timing of the responses, particularly when no episodic memory for the sequence can be accessed. These results complement a series of studies which has shown that PD patients are impaired in motor imagery (Poliakoff, 2013), i.e., when they have to internally initiate action representations - a process similar to the initiation of predictions of external (action) events. However, PD patients are not impaired in action observation (Poliakoff, 2013), as shown for example by the finding that the observation of another agent's actions affects performance of a motor tasks in PD patients just as it does in healthy controls (Albert, Peiris, Cohen, Miall, & Praamstra, 2010).

4.2 Prediction errors and sequential prediction

The proposed role of the basal ganglia in the generation, selection and timing of forward models of probable forthcoming events (Redgrave, Prescott, and Gurney, 1999; Bischoff-Grethe, Crowley, and Arbib, 2003) led us to hypothesise an increased variability at a fine timescale in the segmentation behaviour of PD patients. This hypothesis was supported by the classifier analysis.

However, an alternative account of basal ganglia involvement in action segmentation would also lead to the prediction of increased variability: The Event Segmentation Theory (EST, Zacks and Swallow, 2007; Kurby and Zacks, 2008; Zacks and Sargent, 2010) proposes basal ganglia involvement in signalling prediction errors when unlikely but salient events occur. According to EST, the end of events is signified by prediction errors ('ES prediction errors', hereafter). The underlying theory is that internal forward models of one event become imprecise when the new event begins, which leads to ES prediction errors. EST therefore argues that compromised basal ganglia function leads to disorganised segmentation behaviour (Zacks and Sargent, 2010), as a lack of dopaminergic error signalling prevents the inference that an event boundary has been passed.

In contrast, we would argue that naturalistic events such as actions are usually probabilistically structured (Csibra, 2007; Colder, 2011; Botvinick and Plaut, 2004; Kilner, Friston, and Frith, 2007; Kilner et al., 2004), i.e., that the occurrence of one event makes certain events more probable, while other events are rendered less likely. Accordingly, probable upcoming actions do not constitute a violation of predictions. Moreover, most events are associated with (and thus expected to have) a set approximate duration. Hence, in naturally timed and canonical action sequences such as our action movies, expectations remain usually unviolated.

The understanding that transitions between actions steps are probabilistic or even near-deterministic in character relates to concept of action hierarchies (Botvinick, Niv, and Barto, 2009; Schwartz, 2006; Grafman, 2003; but see Botvinick and Plaut, 2004). An overarching action goal like, e.g., tidying the kitchen, is composed of a series of action components, each with its own goals such as, e.g., clearing away the dishes and tidying the shelves. Again, each of these actions may comprise different subgoals, such as opening the dishwasher, getting a plate out, opening the cupboard, putting the plate into the cupboard, etc... It has not been spelled out yet at which level of this hierarchy dopaminergic ES prediction errors are to be expected. However, experiments that did vary the hierarchical level on which participants had to segment did not report basal ganglia activity for either coarse (high level) or fine grained (low level) segmentation (Zacks et al., 2001).

In the present study, we could establish that PD patients, both ON and OFF medication, show segmentation judgments that are highly similar to controls' judgments and thus seem to rely on the same structured action knowledge. This finding is difficult to reconcile with the proposal that event segmentation has to rely on dopaminergic ES prediction errors. Moreover, while PD patients OFF medication segmented less normatively

if a movie was completely unknown to them, this deviation was not present for the second segmentation instance; this finding speaks against the idea that action segmentation has to rely on intact dopaminergic innervation. Accordingly, we propose that the basal ganglia play a role in the fast generation of timed predictions for probable next sensory states and their evaluation based on the present sensory input.

This account suggests that the probabilistic structure of actions results in the presence of a number of weighted forward models for probable next action steps in the basal ganglia circuits (see Frank, 2006; Frank and Claus, 2006; Frank, Scheres, and Sherman, 2007 for a computational model of weighted forward models in the basal ganglia for goal-directed behaviour). Because the weighing of these probabilities and their generation is dependent on dopaminergic input, PD patients would be compromised in fast decisions on whether a present sensory input (according to the next action step) is in line with, or deviant from, specific forward models.

4.3 **The anatomic specificity of patient data**

Ascribing function to a specific brain area based on data from participants with neurological changes has some limitations; one of many is that the multitude of changes associated with a different neurological conditions make it difficult to ascertain which affected structure is causally relevant for the specific impaired function. Parkinson's disease is associated with changes not only to the basal ganglia, but also to the prefrontal cortex and hippocampus (Brück et al., 2003; Camicioli et al., 2003; Churchyard & Lees, 1997; Emre, 2003; Scatton et al., 1982). While models of basal ganglia and premotor function drove our hypothesis, our results can obviously not discern the changes to which structure underlie the established changes in behaviour. In fact, internally driven prediction of external events and timing of predictions may well rely on interplay of basal ganglia, thalamus and prefrontal/premotor cortex (Lewis et al, 2004; Schönberger, 2013).

4.4 Showing similarity and highlighting differences: the use of classifiers in patient studies

Every study that tests for the ability of patients to perform a task just as well as healthy participants suffers from a conundrum: It is statistically unsound to test for the validity of the null-hypothesis (Cohen, 1994). The present study circumvents this problem by taking a new approach in implementing a classifier analysis. The idea of this classifier analysis is that if the algorithm learns classification from patient data and this classification is then successfully

applied to the data from healthy controls (or vice versa), similarities between the groups has to be considerably high. In fact, in our case it shows that each action movie has a distinct temporal profile of segmentation judgments that makes it different from all other movies. These profiles of the same movie produced by different people were very similar, regardless whether they reflect the behaviour of healthy controls, medicated PD patients, or PD patients off their dopaminergic medication. In the present study, these findings are supported by the correlation analyses that indicate high reliability. The correlation analyses' findings, as well as the segmentation agreement estimation, fall short of the classifier in that they cannot deliver evidence whether what patients do reliably is, in colloquial terms, the same thing healthy controls do reliably. The classifier yields just this distinction.

We believe these very positive results mark classifiers as a valuable tool to investigate hypotheses that propose that patients are not compromised in a given ability. This type of analysis is particularly appropriate for paradigms that provide rich data, for example, behavioural paradigms which assess reaction times, error rates, and subjective judgments (e.g., confidence judgments) for each task, or - perhaps more obviously - studies combining behavioural data and neural recordings. We included classic statistical approaches in the present paper to show that the classical and the novel approach yield similar results. Since the classifier approach is a positive test for the presence of an effect (classification), we suggest that it surpasses the argumentative power of non-significant findings inherent to many inference statistic approaches.

692 Albert, NB, Y Peiris, G Cohen, RC Miall, and P Praamstra (2010). Interference effects from
693 observed movement in Parkinson's disease. J Motor Behav, 42(2), 145-149.
694
695 Alm, PA (2004). Stuttering and the basal ganglia circuits: a critical review of possible relations. Journal of
696 communication disorders 37(4): 325–69.
697
698 Bailey, HR, JM Zacks, DZ Hambrick, RT Zacks, D Head, CA Kurby, and JQ Sargent (2013). Medial temporal
699 lobe volume predicts elders' everyday memory. Psychological science 24(7): 1113–1122.

700 Baldwin, D, A Andersson, J Saffran, and M Meyer (2008). Segmenting dynamic human action via statistical
701 structure. Cognition 106(3): 1382–407.

702 Balleine, BW, M Liljeholm, and SB Ostlund (2009). The integrative function of the basal ganglia in
703 instrumental conditioning . Behavioural Brain Research 199: 43–52.

704 Bernheimer, H, W Birkmayer, O Hornykiewicz, K Jellinger, and F Seitelberger (1973). Brain dopamine and the
705 syndromes of Parkinson and Huntington Clinical, morphological and neurochemical correlations. Journal of the
706 Neurological Sciences 20(4): 415–455.

707 Berns, GS and TJ Sejnowski (1998). A computational model of how the basal ganglia produce sequences.
708 Journal of cognitive neuroscience 10(1): 108–121.

709 Binder, JR, RH Desai, WW Graves, and LL Conant (2009). Where is the semantic system? a critical review and
710 meta-analysis of 120 functional neuroimaging studies. Cerebral Cortex 19(12): 2767–2796.

711 Birkmayer, W and S Wuketich (1976). Time Course of Nigrostriatal Degeneration in Parkinson ' s Disease *.
712 Journal of Neural Transmission 301.

713 Bischoff-Grethe, A, MG Crowley, and MA Arbib (2003). Movement inhibition and next sensory state
714 prediction in the basal ganglia. In The basal ganglia VI, pp. 267–277. Springer.

715 Bornstein, AM and ND Daw (2012). Dissociating hippocampal and striatal contributions to sequential
716 prediction learning. The European journal of neuroscience 35(7): 1011–23.

717 Botvinick, M and DC Plaut (2004). Doing without schema hierarchies: a recurrent connectionist approach to
718 normal and impaired routine sequential action. Psychological review 111(2): 395–429.

719 Botvinick, MM (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior
720 cingulate function. Cognitive, Affective, & Behavioral Neuroscience 7(4): 356–366.

721 Botvinick, MM, Y Niv, and AC Barto (2009). Hierarchically organized behavior and its neural foundations: a
722 reinforcement learning perspective. Cognition 113(3): 262–80.

723 Buxbaum, LJ, K Kyle, M Grossman, and B Coslett (2007). Left inferior parietal representations for skilled
724 hand-object interactions: evidence from stroke and corticobasal degeneration. Cortex 43(3): 411–423.

725 Buxbaum, LJ, KM Kyle, and R Menon (2005). On beyond mirror neurons: internal representations subserving
726 imitation and recognition of skilled object-related actions in humans. Cognitive Brain Research 25(1): 226–239.

727 Brück, A, Kurki, T, Vahlberg, T, and JO Rinne (2004). Hippocampal and prefrontal atrophy in patients with
728 early non-demented Parkinson's disease is related to cognitive impairement. J Neurolo Neurosurg Psychiatry
729 75:1467-1460

730 Camicioli, R, Moore, M M, Kinney, A, Corbridge, E, Glassberg, K, and , J A Kaye (2003). Parkinson's Disease
731 Is Associated with Hippocampal Atrophy. Movement Disorders 18(7): 784-790

732 Cameron, IGM, M Watanabe, G Pari, and DP Munoz (2010). Executive impairment in Parkinson's disease:
733 response automaticity and task switching. Neuropsychologia 48(7): 1948–57.

734 Churchyard, A and A Lees (1997). The relationship between dementia and direct involvement of the
735 hippocampus and amydgala in Parkinson's disease. Neurology 49:1570-1576.

736 Cohen, J (1994). The earth is round (p<. 05). American psychologist 49(12): 997.

Colder, B (2011). Frontiers: Emulation as an integrating principle for cognition. Frontiers in human neuroscience 5.

Cools, R (2006). Dopaminergic modulation of cognitive function - implications for L-DOPA treatment in Parkinson's disease. Neuroscience and biobehavioral reviews 30(1): 1–23.

Cools, R, RA Barker, BJ Sahakian, and TW Robbins (2001). Enhanced or impaired cognitive function in parkinson's disease as a function of dopaminergic medication and task demands. Cerebral Cortex 11(12): 1136–1143.

Cools, R, RA Barker, BJ Sahakian, and TW Robbins (2003). l-Dopa medication remediates cognitive inflexibility, but increases impulsivity in patients with Parkinsonâs disease. Neuropsychologia 41(11): 1431–1441.

Coull, JT, and AC Nobre (2008). Dissociating explicit timing from temporal expectation with fMRI. Current Opinion in Neurobiology 18:132-144

Crawford, T, S Goodrich, L Henderson, and C Kennard (1989). Predictive responses in parkinson's disease: manual keypresses and saccadic eye movements to regular stimulus events. Journal of Neurology, Neurosurgery & Psychiatry 52(9): 1033–1042.

Csibra, G (2007). Action mirroring and action understanding: An alternative account. Sensorimotor foundations of higher cognition: Attention and performance 22: 435–480.

Dagher A, AM Owen, H Boecker, DJ Brooks (2001). The role of the striatum and hippocampus in planning: a PET activation study in Parkinson's disease. Brain 124(Pt 5): 1020–1032.

Daw, ND, SJ Gershman, B Seymour, P Dayan, and RJ Dolan (2011). Model-based influences on humans' choices and striatal prediction errors. Neuron 69(6): 1204–15.

Decety, J, J Grèzes, N Costes, D Perani, M Jeannerod, E Procyk, F Grassi, and F Fazio (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. . Brain 120(10): 1763–1777.

Diniz, PSR, EAD da Silva, and SL Netto (2010). Digital Signal Processing; Cambridge University Press.

Doya, K (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? Neural networks 12(7): 961–974.

Emre, M (2003) Dementia associated with Parkinson's disease. The Lancet Neurology 2(4):229-237.

Flowers, K (1978). Lack of prediction in the motor behaviour of parkinsonism. Brain 101(1): 35–52.

Frank MJ, LC Seeberger, and RC O'Reilly. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306: 1940–43.

Frank, MJ (2006). Hold your horses : A dynamic computational role for the subthalamic nucleus in decision making 19: 1120–1136.

Frank, MJ and ED Claus (2006). Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. Psychological review 113(2): 300–26.

Frank, MJ, A Scheres, and SJ Sherman (2007). Understanding decision-making deficits in neurological conditions: insights from models of natural action selection. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 362(1485): 1641–54.

Grafman, J (2003). The structured event complex and the human prefrontal cortex. Cognitive Processes and Economic Behaviour p. 209.

Graybiel, AM (1998). The basal ganglia and chunking of action repertoires. Neurobiology of learning and memory 70(1): 119–136.

Graybiel, AM (2005). The basal ganglia: learning new tricks and loving it. Current Opinion in Neurobiology 15(6): 638–644.

Hoffman, P, RW Jones, and MAL Ralph (2012). The degraded concept representation system in semantic dementia: damage to pan-modal hub, then visual spoke. Brain 135(12): 3770–3780.

Jay, T (2003). Dopamine: a potential substrate for synaptic plasticity and memory mechanisms. Progress in Neurobiology 69:375-390

Kalbe, E, P Calabrese, N Kohn, R Hilker, O Riedel, HU Wittchen, and J Kessler (2008). Screening for cognitive deficits in Parkinson's disease with the Parkinson neuropsychometric dementia assessment (PANDA) instrument. Parkinsonism & related disorders, 14(2), 93-101.

Kilner, JM, KJ Friston, and CD Frith (2007). The mirror-neuron system: a Bayesian perspective. Neuroreport 18(6): 619–623.

Kilner, JM, C Vargas, S Duval, S-J Blakemore, and A Sirigu (2004). Motor activation prior to observation of a predicted movement. Nature neuroscience 7(12): 1299–301.

Knowlton, BJ, JA Mangels, and LR Squire (1996). A neostriatal habit learning system in humans. Science, 273(5280), 1399-1402.

Kurby, CA, LKE Asiala, and SR Mills (2014). Aging and the segmentation of narrative film. Aging, Neuropsychology, and Cognition 21(4): 444–463.

Kurby, CA and JM Zacks (2008). Segmentation in the perception and memory of events. Trends in cognitive sciences 12(2): 72–9.

Lewis, PA, AM Wing, PA Pope, P Praamstra, and RC Miall (2003). Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisiation, and continuation phases of paced finger tapping. Neuropsychologia 42:1301-1312

Murphy, K. (2012). Machine Learning, a probabilistic perspective; MIT press.

Newtson, D and G Engquist (1976). The perceptual organization of ongoing behavior. Journal of Experimental Social Psychology 12(5): 436–450.

Redgrave, P, TJ Prescott, and K Gurney (1999). The basal ganglia: a vertebrate solution to the selection problem? Neuroscience 89(4): 1009–1023.

Rohenkohl, G., AM Cravo, V Wyart, and AC Nobre (2012). Temporal Expectation Improves the Quality of Sensory Information. The Journal of Neuroscience, 32(24), 8424-8428.

Scatton, C, L Rouquier, F Javoy-Agis, and Y Agid (1982). Dopamine deficiency in the cerebral cortex in Parkinson disease. Neurology 32(9):1039-1040

Schiffer A-M, C Ahlheim, K Ulrichs, and RI Schubotz (2013). Neural changes when actions change: Adaptation of strong and weak expectations. Human Brain Mapping 34(7): 1713–1727.

Schiffer, A.-M., F Waszak, and N Yeung, N. (2015). The role of prediction and outcomes in adaptive cognitive control. Journal of Physiology – Paris. DOI: 10.1016/j.jphysparis.2015.02.001

Schönberger, AR, MT Barbe, K Hagelweide, AB Kühn, GR Fink, and RI Schubotz (2013). Joint principles of motor and cognitive dysfunction in Parkinson's disease. Neuropsychologia, 51(8), 1417-1425.

Schubotz, RI, F M Korb, A-M Schiffer, W Stadler, and DY Von Cramon (2012). The fraction of an action is more than a movement: Neural signatures of event segmentation in fmri. NeuroImage 61(4): 1195–1205.

Schwartz, M (2006). The cognitive neuropsychology of everyday action and planning. Cognitive Neuropsychology 23(1): 202–221.

825 Shohamy, D, CE Myers, J Kalanithi, and MA Gluck (2008). Basal ganglia and dopamine contributions to
826 probabilistic category learning. Neuroscience & Biobehavioral Reviews 32(2): 219–236.

827 Spunt, RP, EB Falk, and MD Lieberman (2010). Dissociable neural systems support retrieval of how and why
828 action knowledge. Psychological Science 21(11): 1593–1598.

829 Stadler, W, RI Schubotz, DY Cramon, A Springer, M Graf, and W Prinz (2011). Predicting and memorizing
830 observed action: Differential premotor cortex involvement. Human brain mapping 32(5): 677–687.

831 Stocco, A, C Lebiere, and JR Anderson (2010). Conditional routing of information to the cortex: a model of the
832 basal ganglia's role in cognitive coordination. Psychological review 117(2): 541–74.

833 Watson, CE and A Chatterjee (2011). The functional neuroanatomy of actions. Neurology 76(16): 1428–1434.

834 Yagüez, L, AGM Canavan, HW Lange, and V Hömberg (1999). Motor learning by imagery is
835 differentially affected in Parkinson's and Huntington's diseases. Behav Brain Res, 102(1-2), 115-
836 127.

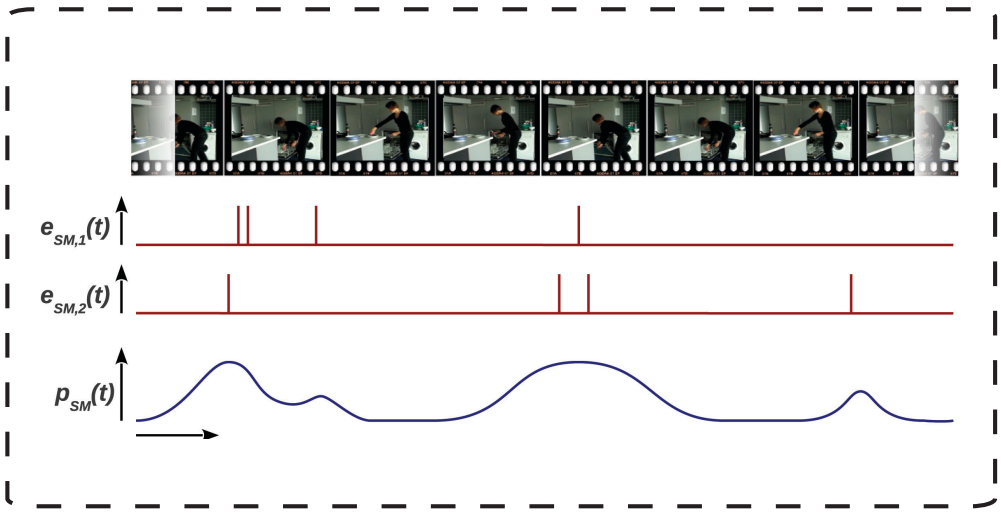
837 Zacks, JM, TS Braver, MA Sheridan, DI Donaldson, AZ Snyder, JM Ollinger, RL Buckner, and ME Raichle
838 (2001). Human brain activity time-locked to perceptual event boundaries. Nature neuroscience 4(6): 651–655.

839 Zacks, JM and JQ Sargent (2010). Event perception: A theory and its application to clinical neuroscience.
840 Psychology of learning and motivation 53: 253–299.

841 Zacks, JM, NK Speer, JM Vettel, and LL Jacoby (2006). Event understanding and memory in healthy aging and
842 dementia of the alzheimer type. Psychology and aging 21(3): 466.

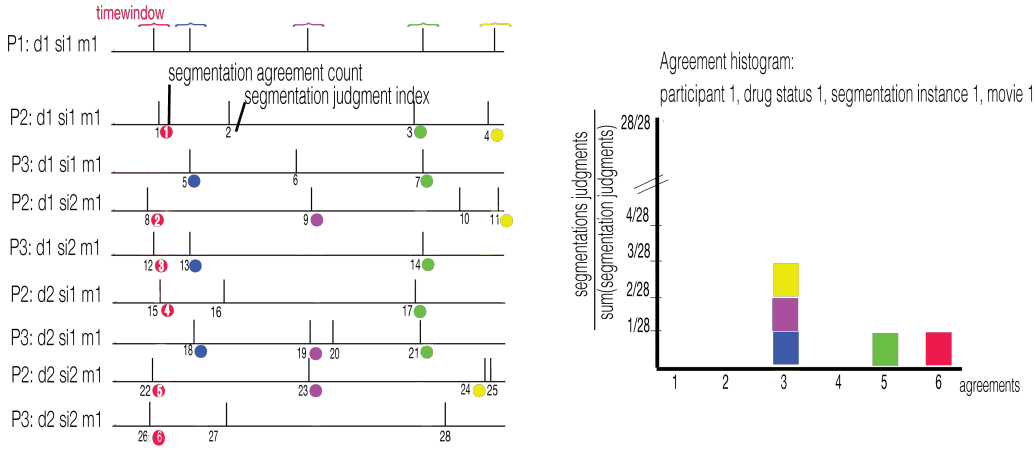
843 Zacks, JM and KM Swallow (2007). Event segmentation. Current Directions in Psychological Science 16(2):
844 80–84.
845
846 Zalla, T, A Sirigu, B Pillon, B Dubois, J Grafman, and Y Agid (1998). Deficit in evaluating pre-determinated
847 sequences of script events in patients with parkinson's disease. Cortex 34(4): 621–627.
848

Figure 1: Probability of segmentation judgments



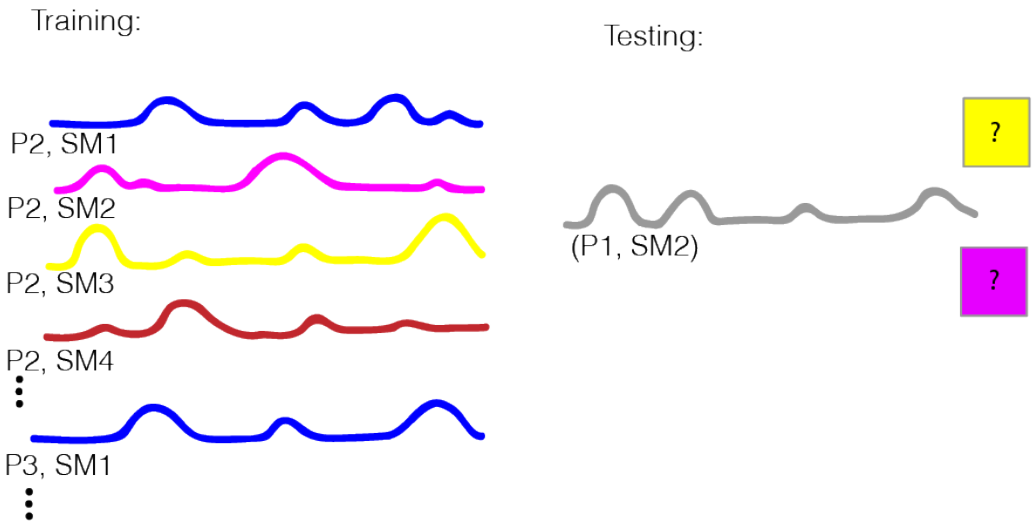
Top row: Frames from an example movie showing an actor clearing out the dishwasher;
2nd & 3rd upper rows: each participant segmented each movie twice (eSM,1 and eSM,2). The red bars correspond to individual segmentation judgments expressed as delta functions, taken from one participant. Each bar represents one segmentation judgment. These delta functions were combined and transformed into temporal patterns representing the probability of a segmentation judgment at each moment in time (probability-density functions), displayed in blue. The classifier analysis used these probability-density functions to predict movie identity.

Figure 2: Schematic representation of segmentation agreement analysis.



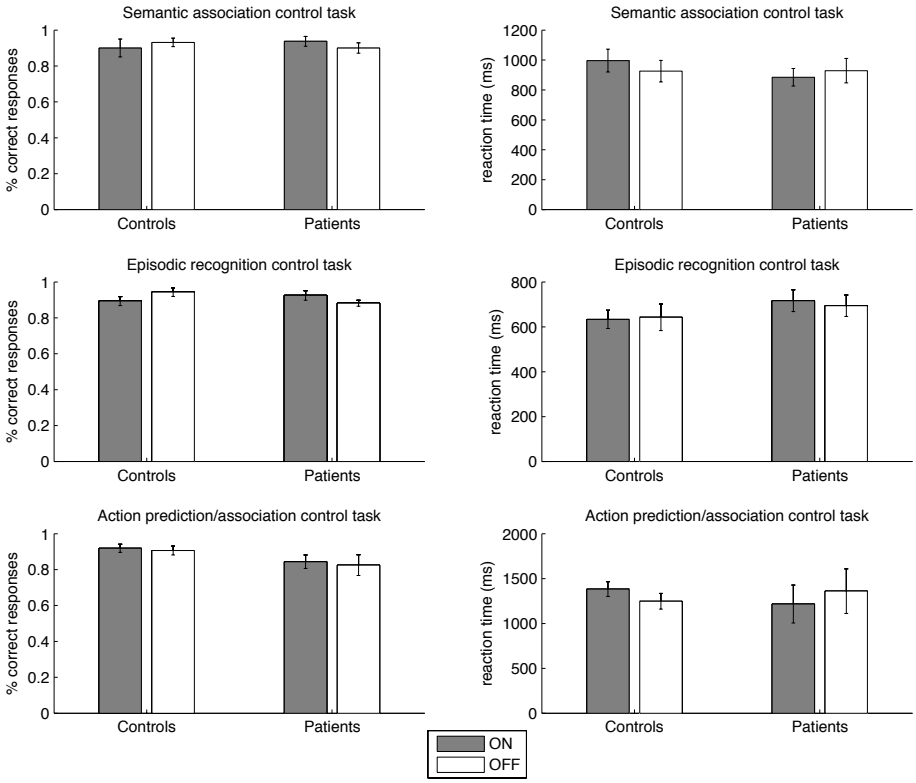
Segmentation agreement scores were calculated for each participant (e.g., P1), under each medication status (here referred to as 'd', or 'drug status', to avoid confusion), for each segmentation instance (e.g., first segmentation, s1) for each movie (e.g., m1). For each segmentation judgment in the respective segmentation instance (left panel, P1: d1 s1 m1), we counted how many other segmentation judgments across the entire group (all participants except the current one and his matched control, in each medication condition, in each segmentation instance, for the *same* movie) would fall into the same time window (e.g., 6 for the first judgments, marked in pink, 3 for the second judgment, marked in purple). For explanation-purposes only, this example assumes a group of 4 participants, instead of the actual 32. This number is then normalised by the overall number of segmentations in the group. This process delivers a histogram of segmentation agreements for each participant in each medication condition, in each segmentation instance, for each movie (displayed on the right). The histogram shows that in this example, one of the segmentation judgments was agreed on in 6 instances (pink) and 3 different segmentation judgments were agreed on 3 times, respectively (purple, magenta, yellow). The combination of these histograms is indicative of the segmentation agreement scores for a subpopulation (eg., PD patients, ON medication, in their first segmentation instance) with the overall group.

Figure 3: Classification on temporal segmentation patterns



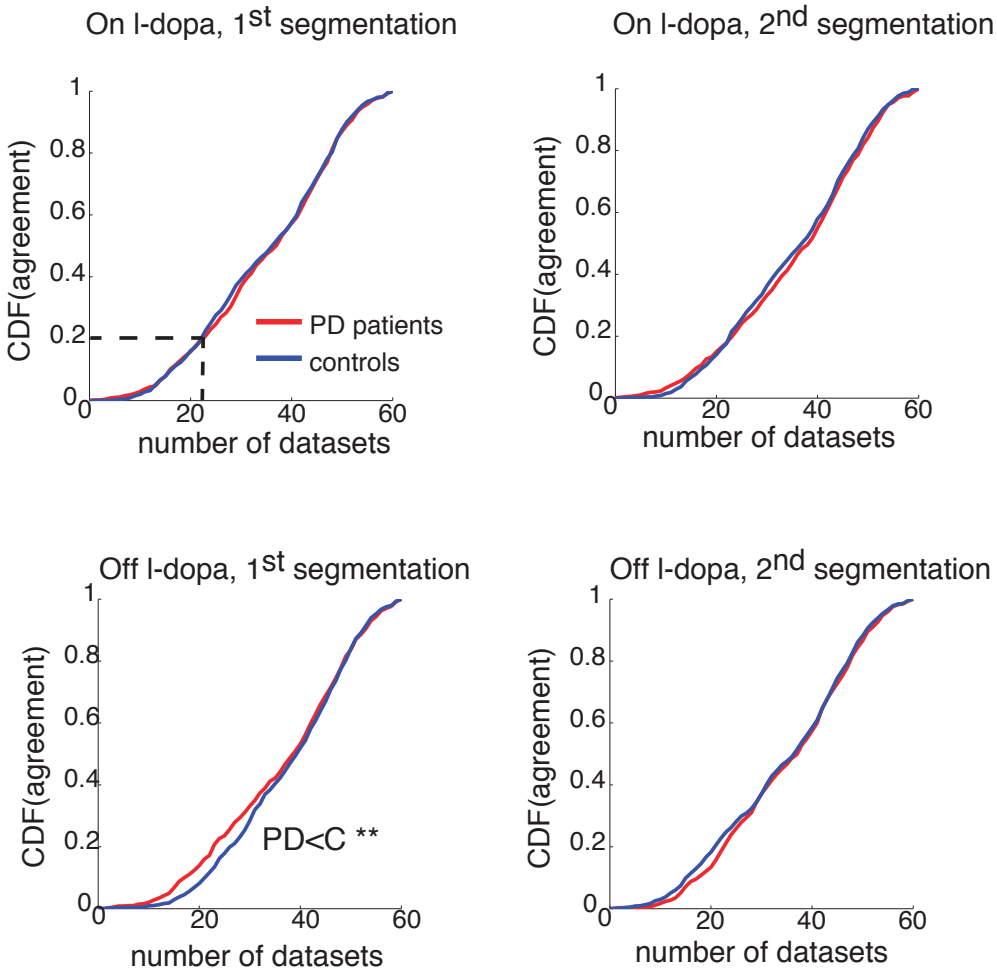
The classifier was trained on a representation of the temporal pattern of responses, i.e., the probability-density functions, capturing the probability of a segmentation judgment over time (see Figure 1), for each movie (SM1, SM2, etc., here limited to 4 movies for presentation purposes only), taken from all participants (P2, P3, etc.) except the one that it was later tested on (P1) and his matched control. In the testing phase, the classifier was iteratively presented with the data from the left-out participant and had to assign one of two possible labels (e.g., doing-the-dishes movie vs sweeping-the-floor movie, here represented as purple and yellow). In the case of across-group classification, the classifier would be presented with the data from the matched control of the left-out participant.

Figure 4: Patients' and controls' performance in the three control tasks.



Legend:
Performance in all control tasks across groups. There were no main effects of group or medication status in any of the tasks.

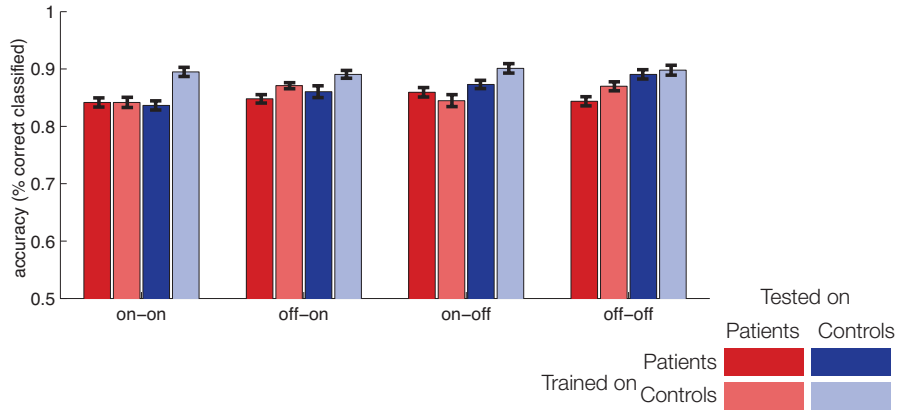
Figure 5: Segmentation agreement across sessions and medication status for PD patients and controls



Legend:
Segmentation agreements, for each participant group (PD – red line/controls – blue line), in each medication status (ON/OFF), for each segmentation instance (first/second). The cumulative distributive function is a random variable, displaying the *area under the curve* calculated from the combined agreement histograms for each group. Considering for example agreement scores in the first segmentation instance ON medication (upper left panel), a probability of agreement of 0.2 is the case in *~23 datasets or less* (see dotted lines) both for PD patients and for healthy controls (the red and blue lines are aligned).

A Kolmogorov-Smirnov showed that the only significant difference was a comparatively lower segmentation agreement for PD patients in the first instance OFF medication (lower left panel), compared to healthy controls in this condition. This deficit is absent during the second segmentation instance in the same session (lower right panel). Additional tests show that this difference is the only statistically significant difference with window sizes varying between 1 and 2.3 seconds. For larger window sizes, all significant differences disappear.

Figure 6: Classifier performance for within and between group classification ON and OFF medication



Legend:
Classifier performance for all tested combinations of training and testing group under all medication conditions. Classification performance for classifiers trained on patients displayed in dark colours, classifier performance for classifiers trained on controls are displayed in lighter colours. Performance of classifiers tested on patients displayed in red and performance of classifiers tested in controls displayed in blue. Medication status in training or testing is indicated by location on the x-axis. on-on: training and testing on medication; off-on: training off, testing on; on-off: training on, testing off; off-off : training and testing off medication. The y-axis starts at 50%, i.e., chance level; error bars show the standard error of the mean. Classifiers tested on controls' data achieve a slightly higher performance (main effect of TESTING GROUP).

Table 1: Descriptive data of patients and healthy controls

	PD: mean - min - max *(STD)*	Controls: mean - min - max *(STD)*
Age (yrs)	61 - 45 - 73 *(7.4)*	61.4 - 51 - 74 *(5.1)*
Edinburgh score	70.3 - 33 - 100 *(17.3)*	73.8 - 50 - 100 *(2.8)*
UPDRS - **ON**	20.9 - 9 - 31 *(6.6)*	1.25 - 0 - 4 *(1.3)*
UPDRS - **OFF**	27.12 - 13 - 37 *(6.9)*	1.1 - 0 - 4 (1.1)
BDI	9.7 - 0 - 19 *(6.1)*	5.8 - 0 - 17 *(4.1)*
PANDA	25.4 - 16 - 30 *(2.8)*	26.1 - 21 - 30 *(2.4)*
Disease duration (yrs)	7 - 2 - 12 *(3.1)*	
Hoehn & Yahr - **ON**	2.4 - 2 - 3 *(0.10)*	
Hoehn & Yahr - **OFF**	2.6 - 2 - 3 *(0.09)*	

Table 2: Condition specific t-values in the comparison of classification performance against chance level (50%). CON: control group, PD: patients

Training-testing group	Training-testing medication	p-value	T-value, all df =15
PD-PD	ON-ON	6*10−16	36
PD-PD	ON-OFF	1*10 −17	46
PD-PD	OFF-ON	1*10−16	40.3
PD-PD	OFF-OFF	3*10−16	37.8
CON-PD	ON-ON	8*10−14	25.7
CON-PD	ON-OFF	1*10−16	39.6
CON-PD	OFF-ON	3*10−17	44.4
CON-PD	OFF-OFF	5*10−19	57.7
PD-CON	ON-ON	1*10−15	33.7
PD-CON	ON-OFF	3*10−18	50.9
PD-CON	OFF-ON	4*10−15	31.3
PD-CON	OFF-OFF	4*10−19	58.8
CON-CON	ON-ON	6*10−16	36
CON-CON	ON-OFF	6*10−19	57.1
CON-CON	OFF-ON	4*10−17	43.3
CON-CON	OFF-OFF	2*10−18	52.4

Table 3: Description of the movies in the segmentation tasks

Movie content	Length
Actor irons shirts, folds onto table.	110 s
Actor finds sugar spilled on floor, takes broom, sweeps floor.	55 s
Actor takes clothes off the line, folds them away.	143 s
Actor clears out the dishwasher and sorts dishes into cupboards.	69 s
Actor gets dressed (coat, boots and scarf), leaves room.	43 s
Actor finds lamp not working, changes light bulb.	53 s
Actor pours milk into cup, spills coffee, gets cloth, wipes table.	44 s
Actor takes a photograph of flowers on a table.	62 s
Actor takes hand pump off bike, starts pumping air into tyre.	76 s
Actor washes and cuts tomatoes, places both into bowl.	143 s
Actor sticks poster to to wall using sellotape.	50 s
Actor cleans dishes by hand.	119 s

Table 4: Overview of individual medication. Dopamine agonists were discontinued up to 36 h (Piribedil: 36 h, Ropinirole/Pramipexole 25 h) and replaced by L-Dopa until complete cessation 14 h before testing.

Patient	Medication
P1	Pramipexole 2,1 mg, L-Dopa 850mg, Selegiline 5mg, Benserazide 75mg, Carbidopa 137,5mg, Entacapone 1000mg
P2	Amantadine 150mg, L-Dopa 600mg, Piribedil 50mg, Entacapone 1000mg, Carbidopa 100mg, Benserazide 25mg
P3	Piribedil 100mg, L-Dopa 300mg, Carbidopa 75mg
P4	Rotigotine 4mg, Rasagiline 1mg
P5	Piribedil 400mg, L-Dopa 400mg, Carbidopa 100mg
P6	L-Dopa 300mg, Carbidopa 75mg, Pramipexole 3,15mg
P7	Pramipexole 2,1mg, Selegiline 5mg
P8	Pramipexole 2,1 mg, Rasagiline 1mg
P9	Pramipexole 2,1mg, Rasagiline 1mg
P10	Pramipexole 2,36, Rasagiline 1mg
P11	Ropinirole 12mg, Rasagiline 1mg
P12	Pramipexole 2,62, L-Dopa 700mg, Benserazide 25mg, Amantadine 300mg, Tolcapone 300mg, Carbidopa 150mg
P13	Amantadine 200mg, L-Dopa 200mg, Benserazide 50mg, Selegiline 10mg
P14	Pramipexole 3,15, Rasagiline 1mg, 225 L-Dopa, Carbidopa 56,25mg, Entacapone 600mg
P15	Ropinirole 2mg, Rasagiline 1mg, Amantadine 200mg
P16	Amantadin 200mg, Rasagiline 1mg, L-Dopa 218,75 mg, Benserazide 43,75mg