Towards a Computational Cognitive Neuropsychology of Wisconsin Card Sorts: A Showcase Study in Parkinson's Disease

Alexander Steinke¹, Florian Lange^{1,2}, Caroline Seer^{1,3}, Bruno Kopp¹

¹Department of Neurology, Hannover Medical School, Carl-Neuberg-Straße 1, 30625, Hannover, Germany,

²Behavioral Engineering Research Group, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium

³Movement Control & Neuroplasticity Research Group, Department of Movement Sciences, KU Leuven, Tervuursevest 101, 3001 Leuven, Belgium

Correspondence concerning this article should be addressed to Alexander Steinke, E-mail: steinke.alexander@mh-hannover.de; Tel.: +49 511 532 3145.

FL received funding from the German National Academic Foundation and BK received funding from the Petermax-Müller Foundation.

The Wisconsin Card Sorting Test (WCST) is an instrument for the clinical assessment of executive functions. Computational modeling of latent cognitive processes offers a route toward improved interpretability of performance on neuropsychological tests. We implemented a computational model by Bishara and colleagues (2010) that allows for evaluating the role of feedback-based attentional learning. We investigated if the model differentiates between Parkinson's disease (PD) patients on and off dopaminergic medication. We reanalyzed data from 32 patients with idiopathic PD and 35 matched healthy controls, which completed a computerized version of the WCST. The PD sample was divided into patients tested on ($n=18$) and off ($n=14$) dopaminergic medication. Model performance was assessed via posterior probabilities and simulations of WCST error scores. Individual model parameters were used for group comparisons. The best performing model configuration showed a single learning rate parameter for positive and negative feedback and recovered the observed WCST perseveration error and set-loss error rates. The occurrence of inference errors could not be accounted for by that model configuration. We did not observe evidence for differences in model parameters between the examined groups of individuals. Our results indicate that assuming distinct, feedback-specific learning processes is not superior to feedback-type unspecific processes in accounting for performance on the computerized WCST. Specifically, the studied model successfully accounted for some, but not all, aspects of test performance. The model parameters did not differentiate between PD patients on and off their dopaminergic medication, a finding that we discuss in the context of study design.

Keywords: computational cognitive neuropsychology, attentional learning, executive function, Wisconsin Card Sorting Test; Parkinson's disease

Towards a Computational Cognitive Neuropsychology of Wisconsin Card Sorts: A Showcase Study in
Parkinson's Disease

The Wisconsin Card Sorting Test (WCST; Berg, 1948; Grant & Berg, 1948; Heaton, Chelune, Talley, Kay, & Curtiss, 1993; Schretlen, 2010) represents a standard tool for the neuropsychological assessment of executive functions (Diamond, 2013; Miller & Cohen, 2001). Performance on the WCST is sensitive to frontal lobe lesions (e.g., Alvarez & Emory, 2006; Barceló & Knight, 2002; Demakis, 2003; Milner, 1963) and to numerous neurological diseases, including Parkinson's Disease (PD; e.g., Kudlicka, Clare, & Hindle, 2011; Dirnberger & Jahanshahi, 2013), or Amyotrophic Lateral Sclerosis (e.g., Lange, Vogts, et al., 2016). However, the complexity of the WCST renders it difficult to infer which cognitive processes generate the observable behavior on the test. In this study, we apply computational modeling as a possible route toward a better interpretability of WCST performance.

Insert Figure 1 about here.

The WCST is considered as an assessment instrument for individual abilities in cognitive flexibility (Diamond, 2013; Miller & Cohen, 2001). Participants are instructed to sort a stimulus according to one of four key cards. Card sorts can accord to one of three viable rules that refer to the color, the shape, or the number of displayed objects. In order to figure out the prevailing rule, the participant has to follow the examiners feedback, as shown in Figure 1. A positive feedback "Repeat" indicates that the selected key card was correct, and that the applied rule should be maintained. A negative feedback "Shift" indicates an incorrectly selected key card, and that the applied rule should be shifted. The prevailing rule changes after a defined number of correct card sorts. WCST error scores provide primary measures of cognitive flexibility (Lange, Seer, & Kopp, 2017; see Figure 1). Standard measures in the literature are perseverative errors (i.e., a failure to shift away from a rule that was previously correct but is no longer correct) and set-loss errors (i.e., a failure to maintain a correct rule). In a recent study, integration errors have been proposed as an additional error measure that can provide incremental information about participants' WCST performance (Lange, Kröger, et al., 2016). Figure 1 gives a trial-by-trial example of the WCST and the operational definition of perseverative errors, integration errors, and set-loss errors (for further information about the WCST, see; Bishara et al., 2010; Eling, Derckx, & Maes, 2008; Lange et al., 2017; Nyhus & Barceló, 2009).

While these three types of WCST errors (perseverative error, integration error, set-loss error) are clearly distinguishable on an operational level, little is known about the latent cognitive processes that lie behind their appearance. To give one example, it has been suggested that executive functioning performance on the WCST reflects general fluid intelligence (Roca et al., 2010). Following this hypothesis, all three types of WCST errors should be correlates of the same cognitive process, i.e. general fluid intelligence. Following Lange et al. (2017), perseverative errors, integration errors, and set-loss errors result from failures of separable cognitive processes, namely set shifting, rule inference, and set maintenance, respectively (see Figure 1). This study represents an attempt to test the separability of these postulated cognitive processes.

The question then emerges how to infer latent cognitive processes from observable behavior on the WCST. There are mainly two approaches to identify process-behavior-relationships: First, model-free cognitive neuropsychology investigates patterns of behavioral impairments as they occur in various groups of patients in order to discover double dissociations as a vehicle toward the identification of separable cognitive modules (Davis, 2010; Shallice, 1988). In general, the model-free approach has not yet provided convincing results, i.e., the identification of separable cognitive modules, in the realm of executive functions. Second, model-based cognitive neuropsychology offers a method to formalize assumptions about cognitive processes, which are assumed to explain patterns of observed behavior in neurological patients (e.g. Botvinick & Plaut, 2004; Cooper & Shallice, 2006; Farrell & Lewandowsky, 2015; Frank & O'Reilly, 2006; Frank, Seeberger, & O'Reilly, 2004; Mavritsaki, Allen, & Humphreys, 2010; Mavritsaki, Heinke, Deco, & Humphreys, 2009; Plaut, McClelland, Seidenberg, & Patterson, 1996). The modeling approach has repeatedly been applied to the WCST (Amos, 2000; Berdia & Metz, 1998; Bishara et al., 2010; Cella et al., 2013; Dehaene & Changeux, 1991; Farreny et al., 2016; Kaplan, Şengör, Gürvit, Genç, & Güzelış, 2006; Kimberg & Farah, 1993; Levine & Prueitt, 1989; Monchi, Taylor, & Dagher, 2000). Our study is based on the model that was proposed by Bishara and colleagues (2010).

Bishara and colleagues (2010) introduced an attentional learning model (ALM) of WCST performance: The ALM is based on the idea of an attention vector, which represents the attentional priorities given to the three viable task rules. The attention vector is updated on a trial-by-trial basis as a function of learning from feedback. A learning rate parameter (or, equivalently, sensitivity-to-feedback parameter, or attention shifting parameter) determines the strength of attention updating. The model

offers two major configurations: A uni-dimensional variant of the ALM assumes a single learning rate parameter (i.e., r parameter; ALM-1) that was originally introduced by Bishara et al. (2010) as a parameter constraint of the ALM (i.e., " $p = r$ "). A two-dimensional variant of the ALM assumes distinct learning rates for positive and negative feedback (i.e., r parameter for positive feedback, and p parameter for negative feedback that was originally referred to as the parameter constraint " p free"; ALM-2). A detailed account of the ALM is given in the methods section. In previous studies, ALM-2 described individual performance on the Heaton-version of the WCST (Heaton et al., 1993) better than ALM-1 did (Bishara et al., 2010). ALM-2 contributed to the dissociation of psychiatric disorders, like in samples of schizophrenia and bipolar disorder patients (Farreny et al., 2016); and substance-dependent individuals and healthy controls (Bishara et al., 2010). Furthermore, ALM-2 was used for analyses of therapeutic effects in schizophrenia (Cella et al., 2013). However, differences in model performance between ALM-1 and ALM-2 were small (Bishara et al., 2010). Importantly, the ALM-1 and the ALM-2 result in different hypotheses about WCST performance.

The ALM-1 offers the more parsimonious account, with only one freely varying learning rate parameter that needs to be estimated. Inter-individual differences and group differences (as a corollary of the presence of particular neuropathological conditions in groups of neurological patients) would be reducible to this learning rate parameter. Of importance, individuals or groups of individuals with low learning rates are expected to perform worse on all error measures that we consider. This holds true despite the fact that the error scores on the WCST can be categorized by the type of feedback that was provided on the previous trial, which was either positive feedback (in case of "Repeat") or negative feedback (in case of "Shift"). Perseverative errors and integration errors occur after negative feedback, whereas set-loss errors follow positive feedback. Following assumptions of the ALM-1, the same learning rate parameter modulates all errors types: Low overall learning rates should lead to an increase in all three error rates, irrespective of the type of feedback that was provided on the previous trial.

ALM-2 allows generating more variable predictions concerning suitable patterns of behavioral observations in individuals and groups of individuals. Given that positive feedback "Repeat" can be equated with "reward" and negative feedback "Shift" can be equated with "punishment", low values in r and in p would predict overall higher error rates on all three error scores on the WCST. Lower r parameter values in the presence of normal p parameter values should lead to selectively increased set-loss error rates because only set-loss errors occur after positive feedback. Lower p parameter values in the

presence of normal r parameter values should lead to increased rates of perseverative errors and integration errors in the presence of preserved set-loss error rates because both perseverative errors and integration errors occur after negative feedback.

Reward-based learning and punishment-based learning may be separable cognitive processes (Frank et al., 2004; Palminteri et al., 2012) that may be subserved by distinct neural processes. Reward-based learning has been mapped to midbrain dopamine systems (Daw & Tobler, 2013; Glimcher, 2011; Schultz, Dayan, & Montague, 1997). More specifically, positive reward prediction errors (i.e., the reception of more reward than expected) are associated with phasic activity bursts of these dopaminergic neurons, whereas negative reward prediction errors (i.e., the reception of less reward than expected) are associated with a phasic suppression of their activity. Numerous studies demonstrated deficits in learning from positive and negative outcomes in dopamine-related diseases, such as PD (Cools, Altamirano, & D'Esposito, 2006; Cools, Barker, Sahakian, & Robbins, 2001; Gotham, Brown, & Marsden, 1988; Swainson et al., 2000; Vaillancourt, Schonfeld, Kwak, Bohnen, & Seidler, 2013) and Gilles de la Tourette's syndrome (Worbe et al., 2011; Palminteri et al., 2009; Palminteri et al., 2011).

PD is a progressive neurodegenerative disorder characterized by affected dopaminergic neurons in the substantia nigra pars compacta, which results in dopamine depletion in the basal ganglia (Dirnberger & Jahanshahi, 2013; Kish, Shannak, & Hornykiewicz, 1988; World Health Organization, 2006). The associated motor symptoms improve under dopaminergic medication (Vaillancourt et al., 2013). It has been shown that dopamine depletion in non-medicated PD led to impaired learning from positive outcomes, but it facilitated learning from negative outcomes. Compared to PD patients off dopaminergic medication, PD patients on dopaminergic medication show a reversed pattern in learning from feedback: Learning from positive outcomes was facilitated, but learning from negative outcomes was impaired. The cited evidence for a dopamine-modulated learning from positive and negative feedback in PD comes from a probabilistic reinforcement learning study that was conducted by Frank and colleagues (2004) and replicated in other studies (Bódi et al., 2009; Cools et al., 2006; Moustafa, Cohen, Sherman, & Frank, 2008; Palminteri et al., 2009; Voon et al., 2010; see also Robbins & Cools, 2014).

The current study addresses two questions: First, we examine whether the assumption of distinct learning rates for positive "Repeat" feedback and for negative "Shift" feedback is supported on a computerized variant of the WCST (cWCST), as originally conjectured by Bishara et al. (2010) for the WCST, by comparing ALM-1 and ALM-2 performances. Second, we compare model parameters from

the winning ALM between PD patients on and off their dopaminergic replacement therapy to see whether reported effects of dopamine on learning from feedback (Frank et al., 2004) were replicable on the cWCST.

Method

The present study reanalyzed data from Lange, Seer, et al. (2016). The authors compared samples of PD patients and healthy controls. Initially, 17 PD patients were tested on and off dopaminergic medication in a within-subject design. However, six out of these PD patients were firstly tested on dopaminergic medication and 11 off dopaminergic medication. This within-subject design does not allow for dissociation of learning effects from effects of dopaminergic medication on WCST performance. Therefore, we analyzed only initial testing sessions in this study, which included PD patients both on and off dopaminergic medication. For comparisons of ALM-1 and ALM-2 performances, we used the pooled sample of healthy controls and PD patients.

Insert Table 1 about here.

Participants

Thirty-two patients with idiopathic PD were considered for analyses. This sample was divided into two sub-samples by their dopaminergic medication while testing: Eighteen patients were tested on their usual medication (on dopaminergic medication) and 14 patients were tested after a withdrawal of dopaminergic medication (off dopaminergic medication; median time since last medication = 14 h; range: 4 – 8120 h). As an indicator of the withdrawal of dopaminergic medication, the mean Unified Parkinson's Disease Rating Scale III (UPDRS III; Fahn et al., 1987; Goetz et al., 2008) was used. The mean UPDRS III score, indexing motor symptoms in PD, rose from 15.43 ($SD = 7.19$) on medication to 24.83 ($SD = 14.09$) after the withdrawal in the off-medication sample, indicating worsened motor symptoms. Mean UPDRS III score in the on-medication sample was 15.61 ($SD = 8.60$). Hoehn and Yahr stages, indicating severity of PD in five stages, where stage 1 has the lowest disability (Hoehn & Yahr, 1967), ranged from 1 to 3 (median = 2) in patients on and off medication. Note that there was no differentiation between levodopa or dopamine agonists in the PD patients' medication. The healthy control group consisted of 35 participants without neurologic or psychiatric diseases who were matched in age, gender, and years of education to the initial PD patient sample. Table 1 gives an overview of participant characteristics. Note

that participants completed further neuropsychological assessment that is not reported in this study (see Lange, Seer, et al., 2016; Seer et al., 2017). All participants were offered 25€ as compensation. The study was approved by the ethics committee of Hannover Medical School (vote number: 6589) and all participants gave informed consent in accordance with the Declaration of Helsinki.

Task and procedure

A computerized version of the Wisconsin Card Sorting Test (Barceló, 2003; Lange, Kröger, et al., 2016; Lange, Seer, Finke, Dengler, & Kopp, 2015; Lange, Seer, et al., 2016) was used. Participants were instructed to match a stimulus card to one of four key cards on each trial. Stimulus cards corresponded in each dimension (i.e., color, shape, and number of displayed objects) to a key card representing the three viable rules of color, shape, and number. The applied rule was identified by the pressed response as none of the 24 stimulus cards shared more than one dimension with the same key card. The four key cards were presented invariantly above the stimulus card and were spatially mapped to the responses. Visual cues served as feedback: Positive feedback (the German word for repeat: “BLEIBEN”) indicated that the response and the applied rule were correct and the rule should be repeated on the next trial. Negative feedback (the German word for shift: “WECHSELN”) indicated that the selected response and the applied rule were incorrect and that the participant should shift the rule (Kopp & Lange, 2013). The stimulus card and key cards were presented simultaneously on each trial. Both remained on screen until a response was detected, which was followed by a white screen presented for 800 ms and a subsequently presented feedback cue for 400 ms. Before the next stimulus card appeared, another 1200 ms of white screen was presented. Rule shifts were unpredictable for participants (Altmann, 2004) and appeared only after at least two correct responses. The average run length, i.e. the number of correct card sorts before the rules shifts, was 3.5. The end of the test was reached after 40 runs or at the request of the participant. The minimum of completed runs was 19.

Participants completed a short practice sequence including five rule runs followed by the study sequence while an electroencephalogram was recorded. Further background neuropsychological assessment was subsequently administered. The cWCST was assessed using Presentation® software on a 24-inch screen (distance to participant 120 cm) and responses were collected with a Cedrus® response pad. The study was conducted in a laboratory of the Hannover Medical School, department of neurology. For further details of the experimental setting, see Lange, Seer, et al. (2016).

First, perseverative errors were analyzed. These errors occur on trials after the rule has shifted, indicated by punishing feedback for executing the previously correct rule. When a participant stays with the now incorrect rule and applies it again, a perseverative error is committed. In addition, set-loss errors were analyzed. These errors occur on trials following a positive feedback when the rule has not changed (Lange, Kröger, et al., 2016). Note that this definition differs from the one of Lange, Seer, et al. (2016), as they counted set-loss errors exclusively on the first trial after the correct rule has been identified or the last trial before the rule changes. In this study, set-loss errors included all trials following a positive feedback before the rule shifts (Lange, Kröger, et al., 2016). Furthermore, integration errors were analyzed. These errors occur after two consecutive trials with different executed rules that were followed by negative feedback. Here, participants can integrate information in order to unambiguously identify the correct rule (Lange, Kröger, et al., 2016). Individual error scores were computed separately for perseverative, integration, and set-loss errors by dividing the number of committed errors by the number of trials with a possible error. On the cWCST, reaction times also serve as a dependent variable. Because the ALM does not consider reaction times, we focused exclusively on error scores.

Computational modeling

The ALM introduced by Bishara et al. (2010) describes individual WCST performance by making one-step-ahead predictions of response probabilities that are computed as a function of attention and feedback. The model considers data on a trial-by-trial basis. Therefore, the model parameters are derived from the raw data like standard scores of the WCST and are not transformations of those. Importantly, trial-by-trial modelling allows evaluating the model's ability to recover individual error scores by simulating WCST runs.

Attention on each rule is represented by attention weights. The relative magnitude of an attention weight indicates the strength of attention on a rule: The higher an attention weight is compared to other rules' attention weights, the higher the probability of executing it. As a basic mechanism, attention weights are updated with regard to the received feedback. A positive feedback raises the attention weight of the applied rule and decreases the attention weights of the remaining rules. Thereby, the probability of executing this rule on the upcoming trial increases. Analogous to this, negative feedback decreases the executed rule's attention weight and increases the attention weights of the remaining rules. On the next trial, the probability of switching the applied rule increases. The ALM offers two major configurations: The ALM-1 has one learning rate parameter r . The ALM-2 has distinct learning rates for positive (i.e., r

parameter) and negative feedback (i.e., p parameter). Learning rate parameters range from 0 to 1 with higher values indicating faster attention updating. In both model configurations, the d parameter represents decision consistency, i.e. whether differences between response probabilities are emphasized or equalized. If $d = 1$, response probabilities are proportional to relative attention weights on color, shape and number. If d is higher than 1, applying the rule with the highest attention weight becomes more likely, or in other words, choices are more deterministic. If d is lower than 1, relative differences between attention weights are more equalized, which results in more similar response probabilities. Lastly, on ambiguous feedback trials, attention updating depends on a parameter f that represents attentional focus. Ambiguous feedback cues at least two rules. For example, as illustrated on Trial $t+1$ of Figure 1, the negative feedback is ambiguous as it indicates attention must shift to color and/or number. In contrast, a positive feedback indicates unambiguously that the applied rule was correct. Attention weights are updated after negative feedback according to the relative attention weights on the indicated rules: the greater the relative attention weight, the stronger it will be updated by the feedback. If $f = 1$, attention update is proportional to the current attention weights. If f is higher than 1, the difference between attention weights is emphasized in updating. If f is lower than 1, differences in attention weights are more equalized in updating (for a detailed account of the model's mathematics see the Appendix and Bishara et al., 2010).

We compared 12 different configurations of the ALM. For both, ALM-1 ($p = r$) and ALM-2 (p free), we derived model configurations which are combinations of free or fixed d and f parameters. Model configurations had an individual decision consistency or a decision consistency that is proportional to relative attention weights (free d or $d = 1$, respectively) and an individual attentional focus, attention updates that have equal strengths for each indicated rule, or attention updates that are proportional to the current attention weights (free f , $f \rightarrow 0$ or $f = 1$, respectively).

The computational model is based on the Heaton version of the WCST (Heaton et al., 1993) but adaptations to other versions are feasible (Bishara et al., 2010). The cWCST used here differs from the Heaton version with regard to the criteria for the end of the test and the run lengths and the set of stimuli (i.e. the Heaton version included ambiguous stimuli). We adapted the algorithm provided by Bishara et al. (2010) to the cWCST considering these criteria. Maximum likelihood estimation with 100 iterations per model configuration and participant was used for model fitting. Default settings from Bishara et al. (2010) were used as parameter boundaries for maximum likelihood estimation.

To compare ALM-1 and ALM-2 performances on the cWCST, we simulated WCST error scores and used an information-theoretic approach. Both were conducted using the pooled sample. For each participant and model configuration, cWCST runs were simulated using individual parameters of attention mechanisms and the framework of the sequential learning model. The simulation was conducted by computing one-step-ahead predictions of participant response probabilities as a function of attention and feedback on the previous trial that accorded to the cWCST design. Note that simulations were based on individual estimated model parameters and not observed responses. Error scores were calculated for each simulated procedure. Because on each trial the response was a random draw, one simulated procedure for the same model configuration and participant could differ in presented stimulus cards, responses, rule shifts, run lengths, and error scores. To reduce this noise, procedures for each model configuration and participant were computed independently 1000 times and error scores were averaged over iterations. As a criterion for model configuration performance, simulated and observed error scores were compared (see Bishara et al., 2010; Burnham & Anderson, 2003).

For model selection, we used an information-theoretic approach that quantifies model configuration's performance considering the trade-off between goodness-of-fit and parsimony, i.e. the number of free variables (see also Burnham & Anderson, 2003; Vandekerckhove, Matzke, & Wagenmakers, 2015). We chose the Bayesian Information Criterion (BIC; Schwarz, 1978) as a measure of performance for two reasons: First, the BIC favours simple models with few free parameters (Vrieze, 2012). In a model with fewer free parameters, maximum likelihood estimation is more precise and leads to greater independence among parameters (Bishara et al., 2010). We transformed the BIC for each participant and model configuration into approximate posterior probabilities, which indicate the individual probability for a model under consideration of the data and the competing models (Burnham & Anderson, 2002; Nathoo & Masson, 2016; Wasserman, 2000). By only considering individual differences in the BIC, posterior probabilities give model performance on an intuitive to interpret and comparable scale between individuals (i.e., a probability ranging from 0 to 1; Burnham & Anderson, 2003; Raftery, 1995; Wagenmakers & Farrell, 2004). We computed posterior probabilities considering either all model configurations or posterior probabilities considering two model configurations for pairwise comparisons. We analyzed whether each model configuration performs better than a random-rule-guessing model (baseline model). Therefore, nested relationships between all model configurations and the baseline model were examined by computing pairwise posterior probabilities for each participant. These were

compared on a group level by focusing on mean posterior probabilities over all participants and on an individual level by analyzing the number of participants that did not favor the baseline model (i.e., a posterior probability higher than .5). Second, performances of all model configurations were compared by computing posterior probabilities between all nested and non-nested model configurations (Bishara et al., 2010). The better performing ALM was used for further analyses: Healthy controls and PD patients on and off dopaminergic medication were compared by ALM parameters and perseverative, integration, and set-loss errors.

Statistical analyses

Statistical analyses were carried out by Bayesian approaches as they appropriately address recent critiques of psychological research (Cumming, 2014; Kruschke & Liddell, 2017). In Bayesian hypothesis testing, Bayes factors (BF; Jeffreys, 1961) state the degree of change from the prior odds to the posterior odds of hypotheses and, thereby, directly quantify hypothesis evidence. Parameters of the posterior distribution provide measures of the magnitude and certainty of effects by its mode and highest density interval, respectively (Kruschke & Liddell, 2017). Model fitting, simulations, and descriptive statistics were carried out by means of R and Bayesian analysis using JASP (JASP team, 2017) and R. In order to facilitate interpretation of BF, they were summarized by categories of evidential strength (Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). In Bayesian hypothesis testing, the null hypothesis implies no effect, i.e. equal means or no correlation, while the alternative hypothesis implies an effect, i.e. differing means or a correlation. Because sample sizes were small in the sub-samples of PD patients, Bayesian Kendall rank correlations were used for analyzing associations between error scores and model parameters. Kendall rank correlations are robust to outliers and they do not assume any form of linear dependency between the variables but rather express dependence in terms of monotonicity (van Doorn, Ly, Marsman, & Wagenmakers, 2018). Bayesian Kendall rank correlations were computed using R following van Doorn et al. (2018). All correlational analyses had a default uniform prior distribution (stretched beta prior width = 1; see van Doorn et al., 2018). Error scores and model parameters were compared between groups by Bayesian independent samples t-tests with a default prior Cauchy distribution centered at 0 (width = 0.707).

Results

Simulations

To identify the best performing ALM on the cWCST, for each model configuration and participant in the pooled sample, model parameters from maximum likelihood estimation were used to simulate cWCST error scores. For a scatterplot of observed and simulated error scores as a function of parameter constraints, see Figure 2. Firstly, there were no obvious differences in error score recovery between ALM-1 and ALM-2. Simulations indicated that separate learning rate parameters for positive and negative feedback were not necessary to recover individual error rates. Fixing $d = 1$ did not seem to worsen the match between observed and simulated error scores. An individual parameter representing decision consistency did not contribute to the recovery of individual error scores either. There were no major differences between model configurations recovery of perseverative and set-loss errors with a free f parameter, a fixed $f = 0$, or a fixed $f = 1$. Simulated integration errors showed little individual variation with $f = 0$ and had the highest variance with $f = 1$. However, the contribution of fixing f to the recovery of individual error scores was not obvious, as differences between model configurations in simulated integration errors were small. The ALM-1 configuration with, $d = 1$ and $f = 1$ differed from all other model configurations, as there was no obvious fit of observed and simulated error scores.

Insert Figure 2 about here.

Posterior probabilities

Pairwise posterior probabilities were computed to analyse whether each model configuration performed better than the baseline model. Here, each model configuration under consideration of data was more likely than the baseline model with mean posterior probabilities ranging from .92 to .99. The lowest mean posterior probability of .92 was found for the ALM-1 configuration with $d = 1$ and $f = 1$. Regarding model selection on participant level, model configurations with a free d parameter performed better for all participants than the baseline model. With $d = 1$, the baseline model was more likely than the compared model configuration for at least one participant. Again, the ALM-1 model configuration with $d = 1$ and $f = 1$ performed worst, as five participants favored the baseline model when directly compared to this configuration. In summary, all considered model configurations performed better than a random-rule-guessing model.

Insert Table 2 about here.

In order to identify the best performing model configuration, all model configurations were compared by posterior probabilities (see Table 2). Altogether, posterior probabilities were higher for ALM-1 than for ALM-2. A feedback-type unspecific learning rate performed better than distinct learning rates for positive and negative feedback-type. Fixing $d = 1$ decreased posterior probabilities, indicating that a free parameter for decision consistency increased model performance. Lastly, model configurations with a free f parameter had lower posterior probabilities than model configurations with a fixed f parameter. Here, model configurations with $f = 1$ had higher posterior probabilities than $f = 0$. An attentional focus that accords to current attention weights outperformed all other model configurations. The best performing model configuration was ALM-1 with an individual decision consistency and an attentional focus that matched current attention weights (i.e. d free $f = 1$). To further parse ALM-1 and ALM-2 performance, we compared best performing model configurations of ALM-1 and ALM-2 that both had a free d parameter and fixed $f = 1$. Pairwise posterior probabilities indicated that the ALM-1 had a posterior probability of .52 when directly compared to the ALM-2. On participant level, the ALM-1 was favored by 55% of the participants. Note that there were no differences between groups as the ALM-1 was favored by 54% of the healthy controls, 50% of PD patients off medication, and 61% of PD patients on medication.

Model selection by posterior probabilities was inconclusive for ALM-1 and ALM-2. However, we selected ALM-1 (with d free and $f = 1$) for two reasons: First, there were no obvious differences in individual error score recovery for ALM-1 and ALM-2 indicating a comparable summary of individual data by both models. Second, ALM-1 configuration had two free parameters (i.e., r and d), whereas ALM-2 had three free parameters (i.e., r , p , and d). The ALM-1 gave a more parsimonious summary of individual WCST performance and was chosen for further analyses. For results of the ALM-2 of the following analyses, see the Appendix.

Insert Table 3 about here.

Correlational analysis

Individual estimates of ALM-1 parameters were used for correlational analysis on the pooled sample. Results of two-sided Bayesian Kendall rank correlations are displayed in Table 3. There was substantial

to strong evidence for negative correlations of the r parameter with perseverative and set-loss errors, respectively. The less efficient attention updating was, the more perseverative and set-loss errors were committed. Evidence for negative correlations of d and perseverative errors was substantial and extreme for a negative correlation with integration and set-loss errors. The less dependent on attention weights choices were (i.e. the lower d), the more integration and set-loss errors were committed.

Insert Table 4 about here.

Group comparisons

Results of Bayesian independent samples t-tests are displayed in Table 4. In line with results reported by Lange, Seer, et al. (2016), comparison of healthy controls and PD patients provided strong evidence for the hypothesis of differing set-loss errors and substantial evidence for the hypothesis of differing perseverative errors. When comparing healthy controls and PD patients, evidence for the hypothesis of differing ALM-1 r parameters was anecdotal. In all other comparisons, evidence was in favor of the hypothesis of equal means.

Discussion

In this study, we evaluated the role of attentional learning from positive and negative feedback on a computerized variant of the WCST via computational modeling. Similar to previous studies (Bishara et al., 2010), direct comparisons of model performances of ALM-1 (uni-dimensional learning rate) and ALM-2 (two-dimensional learning rate) were inconclusive. However, ALM-1 sufficiently recovered individual error scores that were assumed to be specific of positive and negative feedback. We found that assuming distinct, feedback-specific learning processes is not superior to feedback-type unspecific processes in accounting for performance on the computerized WCST. We conclude that cWCST performance could be sufficiently described by one learning rate parameter and a parameter for decision consistency. According to dopamine-modulated learning theories, learning from positive and negative feedback in PD should be modulated by dopaminergic medication. However, we did not find evidence for differences in error scores and ALM-1 parameters as a function of dopaminergic medication in PD patients.

Perseverative and set-loss errors follow negative and positive feedback, respectively. However, the ability of the ALM-1 model to recover both error scores indicates that these error types are driven by feedback-type unspecific processes. The learning rate parameter r was negatively correlated with perseverative and set-loss errors. We conclude that, with regard to Lange et al. (2017), the putative cognitive processes of set shifting (giving rise to perseverative errors) and set maintenance (giving rise to set-loss errors) could be in part accounted for by a uni-dimensional learning rate. In future studies, a learning rate parameter of WCST performance could contribute to the differentiation of clinical samples. However, we did not find evidence for differences in ALM-1 learning rate parameter between healthy controls and PD patients as well as between PD patients on and off dopaminergic medication. Possible reasons are discussed below. Furthermore, as the ALM-1 d parameter was negatively correlated with all three types of WCST errors, decision consistency seems to be a general factor that contributes to cWCST performance. The question emerges how the modelled decision consistency relates to general fluid intelligence, which is supposed to contribute to general WCST performance. Functional neuroimaging studies suggest that a multiple-demand pattern of frontal and parietal activity is associated with tests of fluid intelligence (Duncan, 2010). This multiple-demand system shows activity when complex tasks are learned by separating and assembling its sub-tasks (Duncan, 2010, Duncan & Owen, 2000). With regard to the WCST, the multi-demand system could also be associated with general WCST performance. The relation of the modelled d parameter and the activity of such a multi-demand system could be addressed by imaging studies.

Simulations revealed that all model configurations overestimated integration errors. We found evidence that integration errors are mainly driven by decision consistency. In the implemented computational model, decision consistency was not adaptive to changing task demands, like repeating or inferring the correct rule. In other words, the model did not change between states of high flexibility and high stability within one participant. According to the flexibility–stability hypothesis (e.g. Fröber, Raith, & Dreisbach, 2018; Goschke & Bolte, 2014; Hommel, 2015), inference of the correct rule could depend on highly flexible behavior, whereas the repeated execution of the correct rule could be a state of stable behavior. The ALM’s decision consistency is a compromise, as it should explain behavior in both situations. We suggest an adaptive mechanism of decision consistency that expresses the trade-off between flexibility and stability on a trial-by-trial basis. A different account of inference was proposed by, for example, Wilson and Niv (2012; see also Niv et al., 2015). In order to find the stimulus feature

and response that yields the highest reward, this computational model “tests” specific hypotheses and changes attention through stimulus features. Furthermore, Collins and Koechlin (2012) hypothesize that reinforcement learning cannot explain executive functions alone, but an internal mechanism is necessary as well that switches between task sets as a function of subjective values of outcomes. Such mechanisms could be implemented in the sequential learning model of Bishara and colleagues (2010) to successfully recover individual integration errors. Compared to perseverative and set-loss errors, integration errors had the highest variance in this study, as seen in Figure 2. Therefore, a computational quantification of the processes underlying integration errors could contribute to the differentiation of clinical samples.

Evidence for the initially reported PD impairments (Lange, Seer, et al., 2016) was substantial or strong for differences in perseverative and set-loss errors, respectively. However, there was no evidence for the hypothesis of differences in error scores between PD patients on and off dopaminergic medication. In this study, the comparison of model parameters did not contribute to improving this differentiation. Frank et al. (2004) reported dopamine-modulated effects on learning from positive and negative feedback in PD. Their paradigm incorporated explicit visual feedback (i.e., “Correct!” printed in blue as positive feedback and “Incorrect” printed in red as negative feedback). In this study, we used implicit feedback that was either “Repeat” as positive feedback or “Shift” as negative feedback. Implicit feedback might be insufficient to observe the hypothesized effects of dopaminergic medication on cWCST performance in PD. A crucial feature of explicit feedback used by Frank and colleagues might be the semantic values of “correct” and “incorrect” that are associated with reward and punishment. Other variants of the WCST use verbal feedback (Heaton et al., 1993; Nelson, 1976; Schretlen, 2010). In future research, a new WCST variant with explicit visual feedback could contribute to the differentiation of PD patients by their dopaminergic state. In addition, implicit feedback might hamper the dissociation of attentional learning following positive and negative feedback, i.e., biasing model performance comparisons in favor of ALM-1. Following, ALM-2 might account better for performance on a WCST variant with explicit visual feedback. In contrast to the Heaton-version of the WCST (Heaton et al., 1993) that was used by Bishara et al. (2010), the implemented cWCST has an increased number of trials and runs (i.e. completed rules). However, whether this is sufficient for reliable parameter estimation on the cWCST, could be investigated by simulation studies.

The size of our PD sample might have been too small and a balanced within-subject design could have increased the power of group comparisons. Furthermore, our manipulation of dopaminergic

medication might have been too weak to observe differences on WCST performance or ALM parameters. The withdrawal of dopaminergic medication successfully increased PD motor symptoms, as indicated by the UPDRS III score, but did not affect error scores and model parameters. A possible explanation is that the type of dopaminergic medication was not controlled in PD patients. Dopaminergic medication was either levodopa or dopamine agonists which address different dopaminergic receptors (Brusa et al., 2005; Cox et al., 2015; Dirnberger & Jahanshahi, 2013; Moustafa, Herzallah, & Gluck, 2013; Vaillancourt et al., 2013). For example, Eisenegger et al. (2014) reported that administering the dopamine D2/D3 receptor agonist sulpride impairs choice performance but not learning. In contrast, levodopa leads to a non-specific increase in brain dopamine level (Eisenegger et al., 2014). In this study, the sample size was too small to reasonably categorize Parkinson's disease patients by levodopa or dopamine agonists. Furthermore, the loss of dopaminergic neurons in the substantia nigra pars compacta in *early* PD is not uniform, as indicated by in vivo diffusion magnetic resonance imaging (Du et al., 2012; Vaillancourt et al., 2009). Age, disease duration, and the Hoehn and Yahr stage of PD patients in this study were comparable to preceding studies (Cools et al., 2001; Gotham et al., 1988). However, it is important to be highly selective for the disease stage and participant age in future research.

Conclusion

The WCST is a complex neuropsychological assessment with so-far vague interpretation of performance. However, its sensitivity for neurological diseases and psychiatric disorders was demonstrated in a vast number of studies. Computational cognitive neuropsychology gives a possible route towards increased interpretability of WCST performance as it allows testing explicit mathematical hypotheses about putative cognitive processes. Two standard WCST performance measures – perseverative and set-loss errors – were assumed to be distinct correlates of positive and negative feedback learning. In this study, computational modeling revealed that the same latent parameters could account for both error scores in the cWCST.

Compliance with Ethical Standards

No potential conflict of interest was reported by the authors. The study was approved by the ethics committee of Hannover Medical School (vote number: 6589) and all participants gave informed consent in accordance with the Declaration of Helsinki.

Data availability statement

The datasets used and/or analysed during the current study are available from the corresponding author.

References

- Altmann, E. M. (2004). Advance preparation in rule switching: what work is being done? *Psychological Science*, 15(9), 616–622.
- Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: a meta-analytic review. *Neuropsychology Review*, 16(1), 17–42.
- Amos, A. (2000). A computational model of information processing in the frontal cortex and basal ganglia. *Journal of Cognitive Neuroscience*, 12(3), 505–519.
- Barceló, F. (1999). Electrophysiological evidence of two different types of error in the Wisconsin Card Sorting Test. *Neuroreport*, 10(6), 1299–1303.
- Barceló, F. (2003). The Madrid card sorting test (MCST): A task switching paradigm to study executive attention with event-related potentials. *Brain Research Protocols*, 11(1), 27–37.
- Barceló, F., & Knight, R. T. (2002). Both random and perseverative errors underlie WCST deficits in prefrontal patients. *Neuropsychologia*, 40(3), 349–356.
- Berdia, S., & Metz, J. T. (1998). An artificial neural network stimulating performance of normal subjects and schizophrenics on the Wisconsin card sorting test. *Artificial Intelligence in Medicine*, 13(1–2), 123–138.
- Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, 39(1), 15–22.
- Bishara, A. J., Kruschke, J. K., Stout, J. C., Bechara, A., McCabe, D. P., & Busemeyer, J. R. (2010). Sequential learning models for the Wisconsin Card Sort Task: Assessing processes in substance dependent individuals. *Journal of Mathematical Psychology*, 54(1), 5–13.
- Bódi, N., Kéri, S., Nagy, H., Moustafa, A., Myers, C. E., Daw, N., ... Gluck, M. A. (2009). 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*, 132(9), 2385–2395.
- Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, 111(2), 395–429.
- Brusa, L., Tiraboschi, P., Koch, G., Peppe, A., Pierantozzi, M., Ruggieri, S., & Stanzione, P. (2005). Pergolide effect on cognitive functions in early-mild Parkinson's disease. *Journal of Neural Transmission*, 112(2), 231–237.

- Burnham, K. P., & Anderson, D. R. (2003). *Model selection and multimodel inference: a practical information-theoretic approach* (2nd ed.). New York, NY: Springer.
- Cella, M., Bishara, A. J., Medin, E., Swan, S., Reeder, C., & Wykes, T. (2013). Identifying cognitive remediation change through computational modelling-effects on reinforcement learning in schizophrenia. *Schizophrenia Bulletin*, 40(6), 1422–1432.
- Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biology*, 10(3), e1001293.
- Cools, R., Altamirano, L., & D'Esposito, M. (2006). Reversal learning in Parkinson's disease depends on medication status and outcome valence. *Neuropsychologia*, 44(10), 1663–1673.
- Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2001). Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and rule demands. *Cerebral Cortex*, 11(12), 1136–1143.
- Cooper, R. P., & Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychological Review*, 113(4), 887–916.
- Cox, S. M., Frank, M. J., Larcher, K., Fellows, L. K., Clark, C. A., Leyton, M., & Dagher, A. (2015). Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes. *Neuroimage*, 109, 95–101.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7–29.
- Davies, M. (2010). Double dissociation: Understanding its role in cognitive neuropsychology. *Mind & Language*, 25(5), 500–540.
- Daw, N. D., & Tobler, P. N. (2013). Value learning through reinforcement: the basics of dopamine and reinforcement learning. In P. Glimcher and E. Fehr (Eds.), *Neuroeconomics: Decision making and the brain* (pp. 283–298). San Diego, CA: Elsevier.
- Dehaene, S., & Changeux, J. P. (1991). The Wisconsin Card Sorting Test: Theoretical analysis and modeling in a neuronal network. *Cerebral Cortex*, 1(1), 62–79.
- Demakis, G. J. (2003). A meta-analytic review of the sensitivity of the Wisconsin Card Sorting Test to frontal and lateralized frontal brain damage. *Neuropsychology*, 17(2), 255–264.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168.
- Dirnberger, G., & Jahanshahi, M. (2013). Executive dysfunction in Parkinson's disease: a review. *Journal of Neuropsychology*, 7(2), 193–224.

- Du, G., Lewis, M. M., Sen, S., Wang, J., Shaffer, M. L., Styner, M., ... & Huang, X. (2012). Imaging nigral pathology and clinical progression in Parkinson's disease. *Movement Disorders*, 27(13), 1636–1643.
- Duncan J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*. 14(4):172–179.
- Duncan, J. & Owen, A. M. (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Cognitive Neurosciences*. 23(10), 475–483.
- Eisenegger, C., Naef, M., Linssen, A., Clark, L., Gandamaneni, P. K., Müller, U., & Robbins, T. W. (2014). Role of dopamine D2 receptors in human reinforcement learning. *Neuropsychopharmacology*, 39(10), 2366–2375.
- Eling, P., Derckx, K., & Maes, R. (2008). On the historical and conceptual background of the Wisconsin Card Sorting Test. *Brain and Cognition*, 67(3), 247–253.
- Fahn, S., Elton, R. L., & Members of the UPDRS Development Committee (1987). Unified Parkinson's Disease Rating Scale. In S. Fahn, C. D. Marsden, D. B. Calne, & M. Goldstein (Eds.), *Recent developments in Parkinson's disease* (pp. 153–164) Florham Park, NJ: Macmillan Health Care Information.
- Farrell, S., & Lewandowsky, S. (2015). An introduction to cognitive modeling. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 3–24). New York, NY: Springer.
- Farreny, A., del Rey-Mejías, Á., Escartin, G., Usall, J., Tous, N., Haro, J. M., & Ochoa, S. (2016). Study of positive and negative feedback sensitivity in psychosis using the Wisconsin Card Sorting Test. *Comprehensive Psychiatry*, 68, 119–128.
- Frank, M. J., & O'Reilly, R. C. (2006). A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. *Behavioral Neuroscience*, 120(3), 497–517.
- Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, 306(5703), 1940–1943.
- Fröber, K., Raith, L., & Dreisbach, G. (2018). The dynamic balance between cognitive flexibility and stability: The influence of local changes in reward expectation and global task context on voluntary switch rate. *Psychological Research*, 82(1), 65–77.

- Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *PNAS*, *108*(Supplement 3), 15647–15654.
- Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., ... & Dubois, B. (2008). Movement Disorder Society sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders*, *23*(15), 2129–2170.
- Goschke, T., & Bolte, A. (2014). Emotional modulation of control dilemmas: The role of positive affect, reward, and dopamine in cognitive stability and flexibility. *Neuropsychologia*, *62*, 403–423.
- Gotham, A. M., Brown, R. G., & Marsden, C. D. (1988). 'Frontal' cognitive function in patients with Parkinson's disease 'on' and 'off' levodopa. *Brain*, *111*(2), 299–321.
- Grant, D. A., Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*(4), 404–411.
- Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: revised and expanded*. Odessa, FL: Psychological Assessment Resources Inc.
- Hoehn, M. M., & Yahr, M. D. (1967). Parkinsonism onset, progression, and mortality. *Neurology*, *17*(5), 427–427.
- Hommel, B. (2015). Between persistence and flexibility: The Yin and Yang of action control. In A. J. Elliot (Ed.), *Advances in motivation science* (Vol. 2, pp. 33–67). New York, NY: Elsevier.
- JASP Team (2017). *JASP* (Version 0.8.1.2) [Computer software]. Retrieved from <https://jasp-stats.org/>
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, England: Oxford University Press.
- Kaplan, G. B., Şengör, N. S., Gürvit, H., Genç, İ., & Güzeliş, C. (2006). A composite neural network model for perseveration and distractibility in the Wisconsin card sorting test. *Neural Networks*, *19*(4), 375–387.
- Kimberg, D. Y., & Farah, M. J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex organized behavior. *Journal of Experimental Psychology: General*, *122*(4), 411–428.
- Kish, S. J., Shannak, K., & Hornykiewicz, O. (1988). Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease. *New England Journal of Medicine*, *318*(14), 876–880.

- Kopp, B., & Lange, F. (2013). Electrophysiological indicators of surprise and entropy in dynamic rule-switching environments. *Frontiers in Human Neuroscience*, 7:300.
- Kruschke, J. K., & Liddell, T. M. (2017). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and planning from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25(1), 178–206.
- Kudlicka, A., Clare, L., & Hindle, J. V. (2011). Executive functions in Parkinson's disease: Systematic review and meta-analysis. *Movement Disorders*, 26(13), 2305–2315.
- Lange, F., Kröger, B., Steinke, A., Seer, C., Dengler, R., Kopp, B. (2016). Decomposing card-sorting performance: effects of working memory load and age-related changes. *Neuropsychology*, 30(5), 579–590.
- Lange, F., Seer, C., Finke, M., Dengler, R., Kopp, B. (2015). Dual routes to cortical orienting responses: novelty detection and uncertainty reduction. *Biological Psychology*, 105, 66–71.
- Lange, F., Seer, C., & Kopp, B. (2017). Cognitive flexibility in neurological disorders: Cognitive components and event-related potentials. *Neuroscience & Biobehavioral Reviews*, 83, 496–507.
- Lange, F., Seer, C., Loens, S., Wegner, F., Schrader, C., Dressler, D., ... & Kopp, B. (2016). Neural mechanisms underlying cognitive inflexibility in Parkinson's disease. *Neuropsychologia*, 93, 142–150.
- Lange, F., Vogts, M. B., Seer, C., Fürkötter, S., Abdulla, S., Dengler, R., ... & Petri, S. (2016). Impaired set-shifting in amyotrophic lateral sclerosis: An event-related potential study of executive function. *Neuropsychology*, 30(1), 120–134.
- Levine, D. S., & Prueitt, P. S. (1989). Modeling some effects of frontal lobe damage: Novelty and perseveration. *Neural Networks*, 2(2), 103–116.
- Mavritsaki, E., Allen, H. A., & Humphreys, G. W. (2010). Decomposing the neural mechanisms of visual search through model-based analysis of fMRI: top-down excitation, active ignoring and the use of saliency by the right TPJ. *Neuroimage*, 52(3), 934–946.
- Mavritsaki, E., Heinke, D., Deco, G., & Humphreys, G. W. (2009). Simulating posterior parietal damage in a biologically plausible framework: Neuropsychological tests of the search over time and space model. *Cognitive Neuropsychology*, 26(4), 343–390.

- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24(1), 167–202.
- Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology*, 9(1), 90–100.
- Monchi, O., Taylor, J. G., & Dagher, A. (2000). A neural model of working memory processes in normal subjects, Parkinson's disease and schizophrenia for fMRI design and predictions. *Neural Networks*, 13(8–9), 953–973.
- Moustafa, A. A., Cohen, M. X., Sherman, S. J., & Frank, M. J. (2008). A role for dopamine in temporal decision making and reward maximization in Parkinsonism. *The Journal of Neuroscience*, 28(47), 12294–12304.
- Moustafa, A. A., Herzallah, M. M., & Gluck, M. A. (2013). Dissociating the cognitive effects of levodopa versus dopamine agonists in a neurocomputational model of learning in Parkinson's disease. *Neurodegenerative Diseases*, 11(2), 102–111.
- Nathoo, F. S., & Masson, M. E. (2016). Bayesian alternatives to null-hypothesis significance testing for repeated-measures designs. *Journal of Mathematical Psychology*, 72, 144–157.
- Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, 12(4), 313–324.
- Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *The Journal of Neuroscience*, 35(21), 8145–8157.
- Nyhus, E., & Barceló, F. (2009). The Wisconsin Card Sorting Test and the cognitive assessment of prefrontal executive functions: A critical update. *Brain and Cognition*, 71(3), 437–451.
- Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., ... & Pessiglione, M. (2012). Critical roles for anterior insula and dorsal striatum in punishment-based avoidance learning. *Neuron*, 76(5), 998–1009.
- Palminteri, S., Lebreton, M., Worbe, Y., Grabli, D., Hartmann, A., & Pessiglione, M. (2009). Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes. *PNAS*, 106(45), 19179–19184.
- Palminteri, S., Lebreton, M., Worbe, Y., Hartmann, A., Lehericy, S., Vidailhet, M., ... Pessiglione, M. (2011). Dopamine-dependent reinforcement of motor skill learning: Evidence for Gilles de la Tourette syndrome. *Brain*, 134(8), 2287–2301.

- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56–115.
- Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 111–196). Cambridge, MA: Blackwell.
- Robbins, T. W., & Cools, R. (2014). Cognitive deficits in Parkinson's disease: a cognitive neuroscience perspective. *Movement Disorders*, 29(5), 597–607.
- Roca, M., Parr, A., Thompson, R., Woolgar, A., Torralva, T., Antoun, N., ... & Duncan, J. (2010). Executive function and fluid intelligence after frontal lobe lesions. *Brain*, 133(1), 234–247.
- Schretlen, D.J., 2010. *Modified Wisconsin card sorting test (m-WCST)*. Lutz, FL: Psychological Assessment Resources.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593–1599.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464.
- Seer, C., Lange, F., Loens, S., Wegner, F., Schrader, C., Dressler, D., ... & Kopp, B. (2017). Dopaminergic modulation of performance monitoring in Parkinson's disease: An event-related potential study. *Scientific Reports*, 7:41222.
- Shallice, T. (1988). *From neuropsychology to mental structure*. New York, NY: Cambridge University Press.
- Swainson, R., Rogers, R. D., Sahakian, B. J., Summers, B. A., Polkey, C. E., & Robbins, T. W. (2000). Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: possible adverse effects of dopaminergic medication. *Neuropsychologia*, 38(5), 596–612.
- Vaillancourt, D. E., Schonfeld, D., Kwak, Y., Bohnen, N. I., & Seidler, R. (2013). Dopamine overdose hypothesis: evidence and clinical implications. *Movement Disorders*, 28(14), 1920–1929.
- Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (2015). Model comparison and the principle of parsimony. In J. T. Townsend, & J. R. Busemeyer (Eds.), *Oxford handbook of computational and mathematical psychology* (pp. 300–319). New York, NY: Oxford University Press.
- van Doorn, J., Ly, A., Marsman, M., & Wagenmakers, E.-J. (2018). Bayesian Inference for Kendall's Rank Correlation Coefficient. *The American Statistician*, 1–6.

- Voon, V., Pessiglione, M., Brezing, C., Gallea, C., Fernandez, H. H., Dolan, R. J., & Hallett, M. (2010). Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. *Neuron*, 65(1), 135–142.
- Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, 17(2), 228–243.
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1), 192–196.
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, 100(3), 426–432.
- Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, 44(1), 92–107.
- Wilson, R. C., & Niv, Y. (2012). Inferring relevance in a changing world. *Frontiers in Human Neuroscience*, 5:189.
- Worbe, Y., Palminteri, S., Hartmann, A., Vidailhet, M., Lehericy, S., & Pessiglione, M. (2011). Reinforcement learning and Gilles de la Tourette syndrome: dissociation of clinical phenotypes and pharmacological treatments. *Archives of General Psychiatry*, 68(12), 1257–1266.
- World Health Organization (2006). *Neurological disorders: public health challenges*. Geneva, Switzerland: World Health Organization.

Appendix

Attentional learning models for the Wisconsin Card Sorting Test (2010)

A three-element vector (one for each rule) gives attention weights on trial t :

$$a_t = \begin{pmatrix} a_{color} \\ a_{shape} \\ a_{number} \end{pmatrix} \quad (1)$$

In this study, all attention weights were 1/3 on trial 1. They were updated on each following trial based on the given feedback and the applied rule. The strength of the update of attention weights s_t is computed as a function of current attention weights, feedback, and attentional focus f . More generally, on positive feedback trials, s_t is given by:

$$s_{t,i}|positive = \frac{m_{t,k,i}a_{t,i}^f}{\sum_{h=1}^3 m_{t,k,h}a_{t,h}^f} \quad (2)$$

and on negative feedback trials by:

$$s_{t,i}|negative = \frac{(1-m_{t,k,i})a_{t,i}^f}{\sum_{h=1}^3 [(1-m_{t,k,h})a_{t,h}^f]} \quad (3)$$

$m_{t,k}$ is a three-dimensional vector representing matches between a rule dimension i (i.e. color, shape or number) and the chosen key card k . Let $m_{t,k,i}$ be 1 if the stimulus card matches pile k on rule dimension i and let $m_{t,k,i}$ be 0 if not. For example, on the first displayed Trial in Figure 1, the stimulus card was sorted by its shape, resulting in $m_{t,k,i}$ is 1 for shape and 0 for color and number. The parameter for attentional focus f either emphasizes or equalizes differences in updates of attention weights.

Attention weights for the next trial a_{t+1} are computed by integrating feedback from the current trial. Here, the ratio of information integrated from the previous trial and the current trial is given by individual parameters r and p ranging from 0 to 1. After positive feedback:

$$a_{t+1}|positive = (1 - r) * a_t + r * s_t \quad (4)$$

and after negative feedback:

$$a_{t+1}|negative = (1 - p) * a_t + p * s_t \quad (5)$$

The uni-dimensional attentional learning model (ALM-1) is based on equation four that is used for both feedback types (i.e., positive and negative feedback). The two-dimensional attentional learning model (ALM-2) has distinct learning rates for positive and negative feedback as defined in equations four and five.

The predicted probability of choosing key card k on trial t is then given by:

$$P_{t,k} = \frac{m'_{t,k} a_t^d}{\sum_{j=1}^4 (m_{t,j} a_t^d)} \quad (6)$$

with $m'_{t,k}$ is the transpose of the transpose of $m_{t,k}$. Here, d represents participant's decision consistency that either makes responses more deterministic or random.

Results of ALM-2

The ALM-2 with free d and fixed $f = 1$ was the second best performing model configuration by posterior probabilities and gave a comparable simulation of individual error scores. In the following, we repeated all analyses as conducted by ALM-1 with the best performing model configuration of ALM-2.

Insert Table A1 about here.

Correlational analysis

Individual estimates of ALM-2 parameters were used for correlational analysis on the pooled sample. Results of two-sided Bayesian Kendall rank correlations are displayed in Table A1. There was extreme to substantial evidence for negative correlations of the p and d parameters with all three error scores. The less efficient attention updating was after negative feedback and the less dependent choices were on attention weights (i.e. the lower d), the more errors were committed. Evidence for negative correlations of r and perseverative errors was anecdotal and substantial for a negative correlation with set-loss errors. The less efficient attention updating was after positive feedback, the more perseverative and set-loss errors were committed.

Group comparisons

Results of Bayesian independent samples t-tests are displayed in Table A2. In all comparisons, evidence was in favor of the hypothesis of equal means of model parameters.

Insert Table A2 about here.

Table 1. Participant characteristics.

	Healthy Controls (<i>n</i> = 35)	Parkinson's Disease Patients	
		All (<i>n</i> = 32)	On Dopaminergic Medication (<i>n</i> = 18) Off Dopaminergic Medication (<i>n</i> = 14)
<i>Age (years)</i>	62.23 (10.11)	62.63 (9.59)	62.56 (9.20) 62.71 (10.42)
<i>% Female</i>	57	34	28 43
<i>Education (years)</i>	13.60 (4.06)	14.27 (3.76)	14.78 (4.47) 13.58 (2.49)
<i>Disease Duration (years)</i>		7.78 (5.81)	8.94 (6.38) 6.29 (4.81)

Note. Group means with standard deviation in parentheses. Disease duration was defined as the difference between the year of testing and the year of PD symptom onset indicated by the medical records.

Table 2. Mean posterior probabilities of model configurations as a function of parameter constraints.

		f free	$f = 0$	$f = 1$	Baseline
ALM-2	d free	.0289	.1229	.2106	.0003
(p free)	$d = 1$	.0050	.0310	.0002	
ALM-1	d free	.0480	.0802	.2749	
($p = r$)	$d = 1$	.0238	.1743	.0001	

Note. Posterior probabilities are computed by considering all model configurations. The baseline model was a random-rule-guessing model: On each trial, all rules had the same probability to be executed.

ALM-1 = uni-dimensional attentional learning model; ALM-2 = two-dimensional attentional learning model.

Table 3. Correlations between error scores and ALM-1 model parameters.

	<i>r</i>	<i>d</i>
Perseverative Errors	-.22* [-.37, .05]	-.24* [-.39, -.07]
Integration Errors	-.06▲ [-.22, .10]	-.35*** [-.49, -.18]
Set-Loss Errors	-.26** [-.41, -.09]	-.31*** [-.46, -.14]

Note. Kendall rank correlations coefficients. 95% highest density intervals in parenthesis. *** $BF_{10} > 100$ (extreme evidence for the alternative hypothesis); ** $30 > BF_{10} > 10$ (strong evidence for the alternative hypothesis); * $10 > BF_{10} > 3$ (substantial evidence for the alternative hypothesis); ▲ $1/3 > BF_{10} > 1/10$ (substantial evidence for the null hypothesis).

Table 4. Group comparisons by error scores.

	Group	Mode	95% HDI	BF ₁₀
Perseverative Errors	Healthy Controls	.10	[.07, .12]	3.55
	Parkinson Patients	.16	[.12, .20]	
	On Dopaminergic Medication	.17	[.10, .24]	0.41
	Off Dopaminergic Medication	.14	[.09, .19]	
Integration Errors	Healthy Controls	.33	[.27, .39]	0.34
	Parkinson Patients	.37	[.29, .45]	
	On Dopaminergic Medication	.41	[.29, .52]	0.51
	Off Dopaminergic Medication	.33	[.21, .45]	
Set-Loss Errors	Healthy Controls	.04	[.03, .05]	18.12
	Parkinson Patients	.09	[.06, .12]	
	On Dopaminergic Medication	.09	[.05, .13]	0.36
	Off Dopaminergic Medication	.08	[.03, .13]	
<i>r</i>	Healthy Controls	.999	[.998, 1]	1.31
	Parkinson Patients	.992	[.986, .999]	
	On Dopaminergic Medication	.989	[.977, 1]	0.59
	Off Dopaminergic Medication	.997	[.991, 1]	
<i>d</i>	Healthy Controls	.087	[.056, .119]	0.34
	Parkinson Patients	.072	[.054, .089]	
	On Dopaminergic Medication	.074	[.053, .095]	0.35
	Off Dopaminergic Medication	.069	[.036, .103]	

Note. Modes and highest density intervals of posterior distributions as well as Bayes factors are computed from either the Bayesian independent samples t-test of healthy controls versus PD or PD on versus off dopaminergic medication. HDI = highest density interval; BF₁₀ = Bayes factor for testing the alternative-hypothesis against the null-hypothesis.

Table A1. Correlations between error scores and ALM-2 model parameters.

	<i>r</i>	<i>p</i>	<i>d</i>
Perseverative Errors	-.17* [-.32, 0]	-.53*** [-.66, -.35]	-.24** [-.38, -.07]
Integration Errors	.03▲ [-.13, .19]	-.42*** [-.56, -.24]	-.37*** [-.51, -.20]
Set-Loss Errors	-.21** [-.36, -.04]	-.45*** [-.59, -.28]	-.38*** [-.52, -.21]

Note. Kendall rank correlations coefficients. 95% highest density intervals in parenthesis. The ALM-2 had a free *d* parameter and $f = 1$. *** $\text{BF}_{10} > 100$ (extreme evidence for the alternative hypothesis); ** $10 > \text{BF}_{10} > 3$ (substantial evidence for the alternative hypothesis); * $3 > \text{BF}_{10} > 1$ (anecdotal evidence for the alternative hypothesis); ▲ $1/3 > \text{BF}_{10} > 1/10$ (substantial evidence for the null hypothesis).

Table A2. Group comparisons by error scores and model parameters for the ALM-2.

	Group	<i>Mode</i>	95% HDI	BF ₁₀
<i>r</i>	Healthy Controls	1	[.999, 1]	0.56
	Parkinson Patients	.997	[.994, 1]	
	On Dopaminergic Medication	.996	[.989, 1]	0.57
	Off Dopaminergic Medication	1	[.999, 1]	
<i>p</i>	Healthy Controls	.928	[.860, .995]	0.44
	Parkinson Patients	.855	[.741, .968]	
	On Dopaminergic Medication	.838	[.676, 1]	0.35
	Off Dopaminergic Medication	.876	[.695, 1]	
<i>d</i>	Healthy Controls	.074	[.053, .096]	0.26
	Parkinson Patients	.071	[.052, .089]	
	On Dopaminergic Medication	.073	[.050, .096]	0.35
	Off Dopaminergic Medication	.068	[.034, .102]	

Note. Modes and highest density intervals of posterior distributions as well as Bayes factors are computed from either the Bayesian independent samples t-test of healthy controls versus PD or PD on versus off dopaminergic medication. The ALM-2 had a free *d* parameter and *f* = 1. HDI = highest density interval; BF₁₀ = Bayes factor for testing the alternative-hypothesis against the null-hypothesis.

Fig 1 An exemplary trial-by-trial run on a computerized Wisconsin Card Sorting Test (Barceló, 2003; Kopp & Lange, 2013; Lange, Kröger, et al. 2016; Lange, Seer, Finke, Dengler, & Kopp, 2015; Lange, Seer, et al., 2016). On Trial t , a participant sorts the stimulus card by “shape”. This is indicated by its position beneath the far-left reference card, which matches the stimulus card by showing triangular objects. The feedback “Repeat (this rule)” indicates that the applied rule (i.e., sorting cards by shape) is correct on this trial. However, on Trial $t+1$, applying the same rule again is erroneous. The correct rule has changed, as indicated by the feedback “Shift (away from that rule)”. Note that such errors cannot be avoided on the Wisconsin Card Sorting Test. On Trial $t+2$, a *perseverative error* is committed since the participant fails to shift away from the previously applied rule after having received a feedback “Shift” on the previous trial. On Trial $t+3$, the participant has successfully shifted away from the shape rule, but applies the color rule, which is incorrect on this exemplary trial. This error would be scored as an *efficient error* (for a detailed discussion, see Barceló, 1999; Barceló & Knight, 2002). On Trial $t+4$, the participant can infer the correct sorting rule by integrating the information from previous trials. Sorting by the shape rule and sorting by the color rule were incorrect in the recent past, such that sorting should be conducted according the number rule on this trial. Trial $t+4$ shows the failure of such an integrative rule inference, which is hence coded as an *integration error*. On Trial $t+5$, the participant successfully infers the number rule, as indicated by the feedback “Repeat” on that trial. However, on the next Trial $t+6$, sorting according the number rule is not maintained, such that this trial would be scored as a *set-loss error*.

Fig 2 Scatterplots of observed (x-axis) and simulated error scores (y-axis) as a function of parameter constraints. Dots represent participants. A match with the bisectrix indicates a perfect fit between observed and simulated error scores. For participants in the upper left from the bisectrix, error scores are overestimated, and for participants in the lower right from the bisectrix, error scores are underestimated by simulations. ALM-1 = uni-dimensional attentional learning model; ALM-2 = two-dimensional attentional learning model.