A Neurocomputational Model of Learning to Select Actions

Andrea Caso
(andrea@andreacaso.com)

Richard P. Cooper
(R.Cooper@bbk.ac.uk)

Centre for Cognition, Computation and Modelling
Department of Psychological Sciences, Birkbeck, University of London
Malet Street, London, WC1E 7HX, United Kingdom

Abstract

We present an extension of a schema-based architecture for action selection, where competition between schemas is resolved using a variation of a neuroanatomically detailed model of the basal ganglia. The extended model implements distinct learning mechanisms for cortical schemas and for units within the basal ganglia. We demonstrate the functionality of the proposed mechanisms by applying the model to two classic neuropsychological tasks, the Wisconsin Card Sorting Task (WCST) and the Probabilistic Reversal Learning Task (PRLT). We discuss how the model captures existing behavioural data in neurologically healthy subjects and PD patients and how to overcome its shortcomings.

Keywords: schema theory; basal ganglia; Wisconsin Card Sorting Test, Probabilistic Reversal Task

Introduction

Schema theory is a framework based on the idea that behaviour in many areas depends on abstractions over instances, i.e., schemas. In these abstract terms, schema theory is very general and has been applied to different domains such as memory and motor control. Norman and Shallice (1980) applied the theory in the domain of routine sequential action. Their theory proposes that action schemas work in a cooperative or sequential fashion, but also that they compete with each other for activation.

While schema theory is helpful in representing functional interactions in the action-perception cycle, it is not committed to a specific neural implementation. However, at the neural level the basal ganglia have been proposed as a candidate for resolving competition between schemas in order to carry out action selection (Redgrave et al., 2001). In part this is because of their recurrent connections with the cortex.

In the first part of the paper we present a schema-theoretic model of action selection where competition between motor and/or cognitive schemas is resolved using a variation of a neuroanatomically detailed model of the basal ganglia. We assume that schemas are cortically represented but that schema selection (i.e., selecting one from a set of competing schemas) is facilitated by the basal ganglia. The latter receive multiple signals from the cortex but they are presumably 'content-free'. In other words, unlike their corresponding cortical structures, they are not directly related to the stimulus features. Following the description of the model we propose how learning may occur in the model subsequent to reward, introducing two parameters that drive separate learning mechanisms. Then, we proceed to present two examples of the model applied to two tasks: Wisconsin Card Sorting Task (WCST) and a variant of the Probabilistic Reversal Learning Task (PRLTv). We discuss computational results, the model fit with existing empirical data, and experiments that could further validate the model.

The Extended Schema-Theory Model

At a general level, the model can be understood as two systems or layers of computational units that feed signals to each other – a cortical system and a basal ganglia system. Each unit within the cortical system corresponds to to a schema, and represent a meaningful action or thought. Cortical units are connected with other cortical units and to the basal ganglia (BG) units (Fig. 1). These BG units take input from all cortical units at the same level of abstraction, generate an output signal, and feed it back to the same cortical units. The BG units serve to resolve competition between same-level schemas via the feedback loop between cortical and BG layers. Below, we will introduce two applications of the model – to the Wisconsin Card Sorting Test (WCST), which makes use of two distinct sets of schemas (cognitive and motor schemas) each with their own BG layer, and to a variation of the Probabilistic Reversal Learning task (PRLTv), which makes use of motor schemas only. First we describe the general model more fully.

Computation

Computation is carried out in both the cortical units and in the five nuclei which make up the basal ganglia (Fig. 2; for a complete description of the basal ganglia functional units see Alexander, 1990) according to the equations given below. In all cases, u_i represents the entry signal to the unit, a_i is the result of integration along the time domain, and o_i represents the output of the unit. The function σ computes the sigmoid function of the input, ensuring output values are bounded between 0 and 1. Sigmoid functions have a fixed slope but variable threshold. Varying the threshold of cortical or striatal units alters the way competition between units is carried out, and can be considered a function of phasic dopamine present in the circuit.[1]

[1] In a separate simulation it has been shown that the level of external dopamine from the substantia nigra pars compacta (SNpc) unit can be simulated by varying the threshold of the saturation curve in the striatum (β_{sma}), without making use of an additional unit.

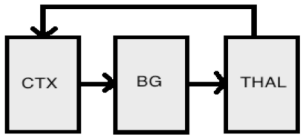

Figure 1: Schematic of the basal ganglia. Legend: Cortex-Thalamic complex (CTX-THAL), Striatum (STR), Subthalamic nucleus (STN), Globus Pallidus Internal/External Segment (GPi and GPe)

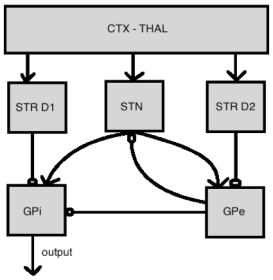

Figure 2: Schematic of the subunits that compose the basal ganglia. Legend: Cortex-Thalamic complex (CTX-THAL), Striatum (STR), Subthalamic nucleus (STN), Globus Pallidus Internal/External Segment (GPi and GPe)

Action selection is not the product of higher order schemas alone. Environmental features directly excite lower order schemas and can lead to selection of those schemas in the absence of higher order control. An excessive ratio or difference between bottom-up and top-down excitation of the lower-level schemas produces behaviours akin to those seen in some frontal patients (Cooper & Shallice, 2000).

Cortical Units (Motor or Cognitive)

$$u_i \longleftarrow o_{ext,i} + o_{thal,i}$$

$$a_i(t) \longleftarrow \partial \cdot a_i(t-1) + (1-\partial)u_i(t-1)$$

$$o_i \longleftarrow \sigma\,(a_i)$$

Striatum (D1 and D2):

$$u_i \longleftarrow o_{sma,i}$$

$$a_i(t) \longleftarrow \partial \cdot a_i(t-1) + (1-\partial)u_i(t-1)$$

$$o_i \longleftarrow \sigma\left(a_{strD1/D2,i}\right)$$

Subthalamic nucleus:

$$u_{stn,i}(t) \longleftarrow w_{stn}o_{sma,i} + w_{gpe\_stn}o_{gpe,i}\,(t-1)$$

$$a_{stn,i}(t) \longleftarrow \partial \cdot a_{stn,i}(t-1) + (1-\partial)u_{stn,i}(t-1)$$

$$o_{stn,i} \Longleftarrow \sigma\left(a_{stn,i}\right)$$

Globus Pallidus (External Segment):

$$u_{gpe,i} \Longleftarrow w_{stn\_gpe}\sum_i o_{stn,i} + w_{strD2\_gpe}o_{strD2,i}$$

$$a_{gpe,i}(t) \Longleftarrow \partial \cdot a_{gpe,i}(t-1) + (1-\partial)u_{gpe,i}(t-1)$$

$$o_{gpe,i} \Longleftarrow \sigma\left(a_{gpe,i}\right)$$

Globus Pallidus (Internal Segment):

$$u_{gpi,i}(t) \longleftarrow w_{stn\_gpi}\sum_i o_{stn,i} + w_{gpe\_gpi}o_{gpe,i}\,(t-1) + w_{strD1\_gpi}o_{strD1,i}\,(t-1)$$

$$a_{gpi,i}(t) \longleftarrow \partial \cdot a_{gpi,i}(t-1) + (1-\partial)u_{gpi,i}(t-1)$$

$$o_{gpi,i} \longleftarrow \sigma\left(a_{gpi,i}\right)$$

Thalamus:

$$u_i \longleftarrow o_{gpi,i}$$

$$a_i(t) \longleftarrow \partial \cdot a_i(t-1) + (1-\partial)u_i(t-1)$$

$$o_i \longleftarrow -\sigma\,(a_i)$$

Cortical and Basal Learning Mechanisms

The general model includes weighted connections from cortical schema units to basal ganglia units, and weighted connections from basal ganglia units back to cortical units. We assume that the weights are learned by separate reward-based mechanisms (for reasons given below). When the system is provided with positive ($r_i = +1$) or negative ($r_i = -1$) feedback after a response, two separate mechanisms control how the system adapts to new stimuli. We assume the following teaching signals are produced by rewards and activations:

$$R_i = (r_i - a_i) \qquad (1)$$

$$S_i = r_i - \sum_{t=1}^{T-1} 2^{i-T+1} \cdot r_t \qquad (2)$$

In Eq. 1 r_i represents the reward assigned to the i^{th} schema and a_i represents the activation of the i^{th} schema, t represents the trial and T is the total number of trials. Eq. 2 encodes the 'surprise' of the reward and assigns a greater value to the most recent trials, effectively implementing a form of 'memory'.

The teaching signals produce a variation in the threshold of the schema and basal ganglia unit saturation curves, β_{ctx} and β_{str}, respectively, as given by Eq. 4 and 5. Uniformly distributed noise ζ in the range [-0.1,0.1] is also added to prevent deadlock.

$$\beta_{str,i} \longleftarrow \eta_0(\beta_{str,i} - \epsilon_{str} R_i + \zeta) \quad (4)$$

$$\beta_{ctx,i} \longleftarrow \eta_0(\beta_{ctx,0} - \epsilon_{ctx} S_i + \zeta) \quad (5)$$

$$\eta_0(x) = \begin{cases} 1, & x > 1 \\ x, & 0 < x < 1 \\ 0, & x < 0 \end{cases} \quad (6)$$

The left arrow indicates assignment[2]. Eq. 4 describes the change of threshold of the saturation curve of BG units following reward. Decreasing β_{str} augments the probability of the i^{th} schema being selected. Eq. 5 describes the change of threshold of the saturation curve of cortical units following reward. Unlike Eq. 4, the value of the β_{ctx} is centred around $\beta_{ctx,0}$ (set to 0.5 in all simulations). Eq. 6 is a limiting function which ensures that the thresholds remain within range.

Overall, this set of equations attempts to capture the division of labour between cortical structures and the basal ganglia. The two distinct learning signals that drive the overall model behaviour represent the direct (mesocortical, through the ventral tegmental area) and indirect (nigrostriatal, from the substantia nigra pars compacta) influence of dopamine to the task representation in the frontal circuits. Both equations are a function of reward, but while Eq. 4 slowly alters the probability of a channel to being selected, Eq. 5 energises schemas when surprise (the difference between expected and given reward) is high and therefore promotes fast dishabituation. Cognitive control emerges from the interaction between the two mechanisms

Theoretical Commitments

The core theoretical commitments of the model are the presence of cortical schemas, the presence of the basal ganglia that act as a content-free action selection device, and two different learning mechanisms for cortical schemas and the basal ganglia. Provided that the learning functions are both based on reward, the analytical form of the functions constitute peripheral hypotheses. Other peripheral hypotheses include the value of the threshold above which a schema is considered selected and the task-dependent number of schemas. The model can also be extended to accommodate other kinds of computation, such as that carried out in the cerebellum.

Model Applied to the WCST

Task and model description

In the Wisconsin Card Sorting Task (WCST), participants are required to sort a series of cards into four categories based on binary (i.e., correct/incorrect) feedback (Heaton, 1981). Each card shows one, two, three or four shapes, printed in one of four colours, and there are four shapes (triangle, star, cross, circle). It is therefore possible to sort cards according to colour, number or shape. To succeed, participants must match each successive card with one of four target cards (which show One Red Triangle, Two Green Stars, Three Yellow Crosses, Four Blue Circles), and use the subsequent feedback to discover the appropriate rule. However, once they have discovered the rule (as indicated by a succession of 10 correct sorts), the experimenter changes the rule without notice. The task yields a number of dependent measures, including the number of rules obtained (with a deck of fixed size – typically 64 or 128 cards), the number of cards correctly sorted, the number of perseverative errors (i.e., errors where the participant persists in using a rule despite having received negative feedback) and the number of Set Loss errors (i.e., errors where the participant fails to stick with a rule despite positive feedback).

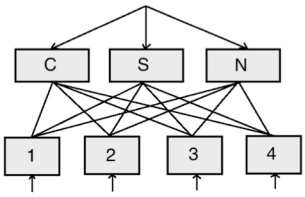

Figure 3: Schematic of the model, not showing competition between schemas. Cognitive schemas (top row) send signals to the motor schemas (bottom row).

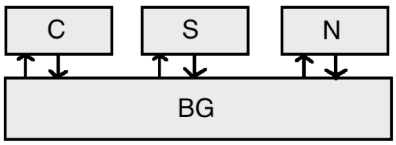

Figure 4: Schematic of the competition between schemas. The basal ganglia units compute the amount of inhibition that each schema receives given the activation of the others. Only cognitive schemas are shown here.

The model comprises three cognitive schemas and four motor schemas (see Fig. 3).[3] Cognitive schemas represent the selection rules (Sort by Colour, Sort by Number, Sort by Shape) while the four motor schemas represent the acts of putting the stimulus card below each of the four target cards. All schemas send signals to the basal ganglia units at the same level of hierarchy (Fig. 4), but only cognitive schemas implement the learning mechanisms outlined in Eq.4-6. Each schema has an activation level that varies over time as a function of input from various sources. Motor schemas are fed by cognitive schemas, and the signal from the cognitive layer to motor layer is rule-dependent. If, for instance, the stimulus card displays three red circles, the shape schema

[2] In assignment the value at the current trial is equal to a function of the same variable in the previous trial. Initial values are 0.5 plus a minimal amount of noise to randomise the first response.

[3] Source code for the simulation, including a complete list of parameters and their values, is available from the first author on request.

will excite the fourth motor schema (Four Blue Circles), the number schema will excite the third motor schema (Three Yellow Crosses), and the colour schema will excite the first motor schema (One Red Triangle).

Motor schemas are also fed by environmental cues which depend on the stimulus card features. Thus, when cognitive schemas are not strong enough to influence motor schemas, stimulus features alone may drive action selection. Feedback is given after each trial, and it drives learning within the cognitive schemas and their BG units as outlined in the previous section (Eq. 4 and 5). Learning in the motor schemas and their associated BG units is unnecessary in the WCST because randomisation of stimuli prevents a preference for a card position from being formed. A typical run of the task is shown in Fig. 5.

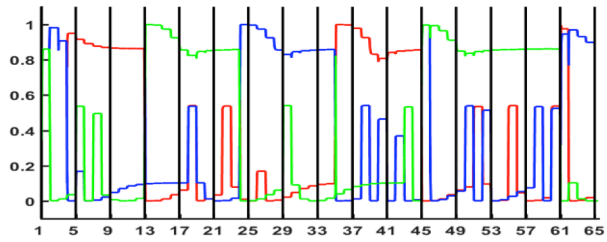

Figure 5: Cognitive schema activation in a typical run of the WCST. The red, green and blue lines represent the colour shape and number schemas, respectively. Black vertical lines have been plotted every 4 trials.

Simulation and results

We simulated 20 subjects for each value of the learning rates ε_{ctx} and ε_{str}, for a total of 560 subjects and recorded the relevant dependant variables (Fig. 6). Total Errors (TE), Perseverative Errors (PE) and Non Perseverative Errors (NPE) are all monotonic functions of ε_{str} and ε_{ctx} while Set Loss (SL) errors show a more erratic pattern. The value of the analysed dependent variables is a function of both ε_{str} and $\varepsilon_{ctx,}$ but also of external activation of cognitive and motor schemas. These signals act as modulators between internal and external attentional process. An excessively low/high value of cognitive/motor external activation signals produces a general increase in all kind of errors. Varying these parameters produces performance more similar to behaviour exhibited by some frontal patients, where environmental cues drive action selection (Cooper & Shallice, 2000). Once baseline values for external excitations are set, we observe how the values of dependent variables fit data from young, older adults, and Parkinson's Disease (PD) patients. PD results from reduced dopaminergic input to the striatum (Siegelbaum et al., 2000) and it is therefore appropriately modelled by lower values of ε_{str}.

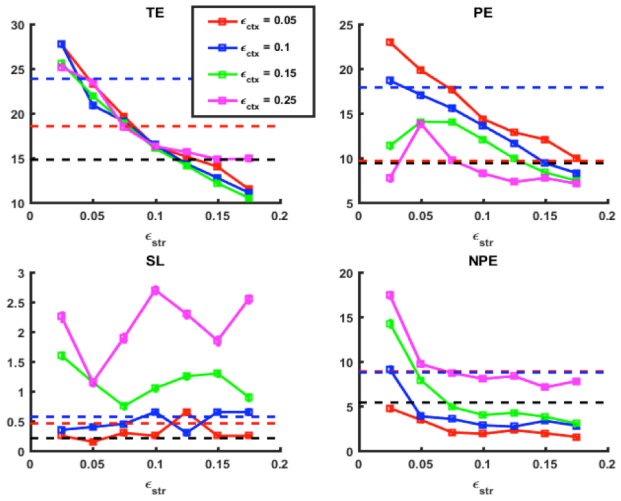

Figure 6: Plot of WSCT simulation results. Dependent variables shown are Total Errors (TE), Perseverative Errors (PE), Set Loss Errors (SL) and Non-Perseverative Errors (NPE). The dashed horizontal black lines, the red lines, and the blue lines represent the mean values of the dependent variables for young participants, older participants, and PD patients, respectively.

Total and Perseverative Errors Empirical data from PD patients (Paolo et al., 1996) performing the WCST show that perseverative errors are significantly greater in non-demented PD patients than in older controls, while the difference between older controls and younger subjects is not significant. The model successfully simulates this pattern of Total Errors and Perseverative Errors in healthy and PD patients with a set of values for (ε_{str} , ε_{ctx}) of (0.15, 0.08) and (0.05, 0.01), respectively. Thus, consistent with the neurophysiological hypothesis, PD patient performance may be accounted for by lower values of ε_{str}.

Set Loss and Non-Perseverative Errors Set loss errors have a different profile from all the other errors, suggesting the presence of distinct cognitive mechanisms underlying these and other errors. Empirical data from young, older controls and PD patients (Paolo et al., 1996) show that SL errors are not significantly greater in non-demented PD patients than in older controls. Paolo et al. (1996) also report that older controls tend to produce more SL errors than younger participants but the difference does not reach significance ($t(89)=1.89$, $p = .062$).

The model does not adequately capture the prevalence of set loss errors, but this limitation might be overcome by choosing parameters more carefully. In addition, it is necessary to further analyse how these errors arise in both the model and in experimental data. SL errors are relatively rare, and do not occur in all attempts at the task (either in human participants or in the model). Further work is required to see whether a more sensitive measurement of SL errors is needed.

Discussion

Simulating the WCST yields an adequate fit with empirical data from healthy young controls and PD patients and it explains how perseveration errors might arise from an

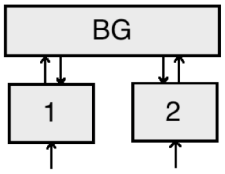

Figure 7: Model diagram for the PRLTv. Unlike the WCST, there are no higher order schemas that control the two lower order schemas.

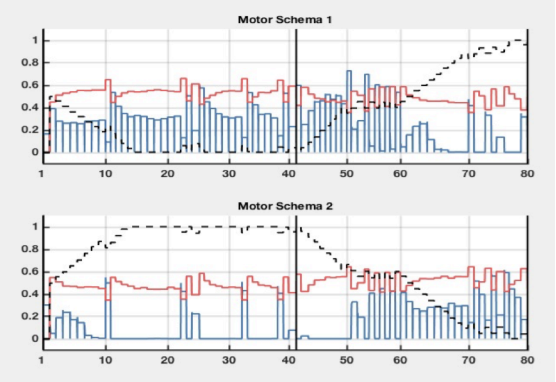

Figure 8: Typical run of the PRLTv. The blue line represents the schema activation while the red line and the dashed black line represent β_{ctx} and β_{str}, respectively.

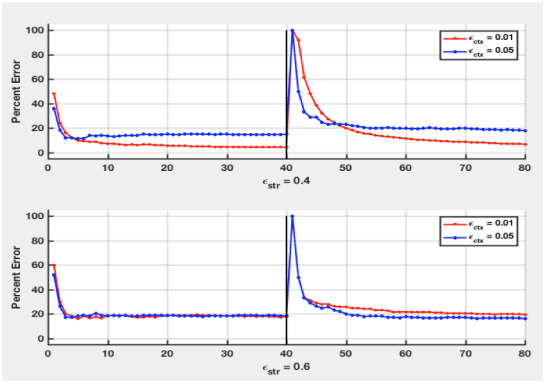

Figure 9: Plots of the model performance in PRLTv for different values of ε_{str} and ε_{ctx} across all trials. Points represent the error percentage for each stage of the task (acquisition and reversal)

impaired selection mechanism, in which rewards do not update quickly enough, or from an impaired schema activation mechanism, where surprising results are not powerful enough to trigger quick selection of a new rule. The dissociation between Set Loss and Perseverative Errors, which reflects the dissociation between distractibility and perseveration (Kaplan et al., 2006), is also replicated. Nevertheless, the model fails to fully explain the difference in Set Loss and Non-Perseverative Errors in healthy and PD populations. It is also unclear whether the difference between young and older control can be modeled with the two learning parameters alone (on the assumption that the trend reported by Paolo et al., 1996, indicates a real effect).

Model Applied to the PRLTv

Task and model description

Here, we apply the general model to a variant of the Probabilistic Reversal Learning Task (PRLTv; Cools et al., 2002). In this task, two stimuli are presented on each trial, but only one is the correct one. However, feedback is unreliable – the subject receives feedback that is correct only 80% of the time. After 40 trials the stimulus that receives the reward (i.e., positive feedback) is reversed. Again, feedback is correct 80% of the time. In the version of the task modelled here (unlike the standard experimental task), we assume that the subject is not told that feedback will be probabilistic. This allows us to test only stimulus-reward contingencies in absence of any super-ordinate rule.

To succeed at the task, subjects have to be able to stick to the first rewarded stimulus despite spurious feedback, but they also have to be able to reverse the choice and not perseverate when the contingency changes. The task is modelled as a simple stimulus-reward association, without higher order rules controlling the selection of lower schemas. The structure of the PRLTv thus is simpler than the one used for the WCST, and consists of only two cortical schemas with their associated basal ganglia units (Fig. 7). A typical run of the model is shown in Fig 8.

Simulation and results

We simulated 25 subjects for two values each of ε_{ctx} and ε_{str} for a total of 100 subjects and display the percentage error across the 80 trials (Fig. 9). Two performance measures are calculated: Errors to Criterion (ETC) are evaluated by counting the number of trials the subject takes to score 8 consecutive correct responses (ignoring spurious feedback). Consecutive-Perseverative (CP) errors are evaluated by counting how many trials from the reversal trial (41st trial) the subject takes to select the correct new response. Both variables are non-normally distributed, and therefore the Kruskal-Wallis H statistic has been used to test differences among the groups.

Errors To Criterion In the acquisition stage, ETCs are not significantly different, irrespective of the parameters (Fig. 10.). On the reversal stage, increasing ε_{str} from 0.4 to 0.6 inverts the ETC trend in the function of ε_{ctx}. The difference in ECT is significant in both the low ε_{str} value *($H(1) = 4.10$, $p = 0.043$)* and the high ε_{str} value *($H(1) = 5.56$, $p = 0.018$)*.

Consecutive Perseverative A low value of ε_{str} generally impairs the model by increasing perseveration (CP = 2), but only for lower values of ε_{str} *($H(1) = 11.68$, $p < 0.001$)* (Fig.11).

Discussion

In the standard version of the Probabilistic Reversal Learning Task (e.g., Swainston et al., 2000), for which data from PD and age-matched controls is available, subjects are encouraged to stick with a rule even if it is occasionally

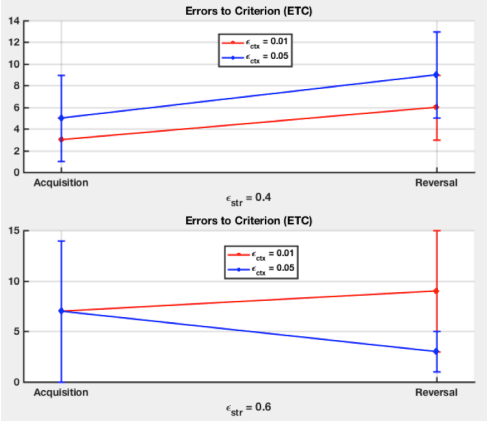

Figure 10: Errors to Criterion (ETCs) are shown here. Points and error bars represent medians and median absolute deviations.

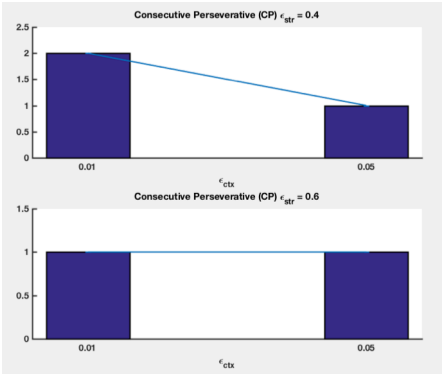

Figure 11: Consecutive Perseverative errors with four different settings (ε_{ctx} ,ε_{str}). Points and error bars represent medians and median absolute deviations.

wrong. This effectively creates a high-level schema. The variant of the task considered here deliberately avoids this and constitutes the lower-level version of the WCST, where only low-level schemas (those schemas that receive direct excitation from the environment) are activated and acted upon by the learning mechanisms. However, because of this difference in task instructions the model cannot be evaluated against the available data. The above results are therefore predictions that remain to be evaluated by contrasting the performance of PD patients and age-matched controls). Our model aims to capture computationally how a simple stimulus-reward association changes in terms of learning mechanisms that act directly on lower level schemas. Therefore the model needs to be experimentally validated with the adjusted behavioural task.

General Discussion

The general model is successful in replicating several empirical results and in reflecting the dissociation between distractibility (exemplified by SL errors in the WSCT and ETC in the PRLTv) and perseveration (exemplified by PE in the WSCT and ETC in the PRLTv). Limitations in accounting for experimental data in the WCST may be overcome by studying how subjects produce NPE and SL errors and whether the model accurately reflects this. Conversely, matching experimental data in the PRLTv requires running new experiments where instructions are reduced to a minimum. Ultimately, the model's purpose is to bridge the concept of neurotransmission, that acts as a medium to increase computational power, and the meaningful unit of action or thought. Thus, while the theoretical core assumptions seem to be capable of reproducing at least two tasks adequately, peripheral hypotheses on the learning mechanisms may require revision to achieve a better fit and to strengthen the link with the neurobiology.

References

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends in Neurosciences*, *13*(7), 266-271.

Caso, A., & Cooper, R. (2017). A model of cognitive control in the Wisconsin card sorting test: integrating schema theory and basal ganglia function. *Cognitive Science Society*.

Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, *17*(4), 297-338.

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using even t-related functional magnetic resonance imaging. *Journal of Neuroscience*, *22*(11), 4563-4567.

Heaton, R. K. (1981). *A manual for the Wisconsin card sorting test*. Western Psychological Services.

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.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior: In RJ Davidson, GE Schwartz, & D. Shapiro (Eds.), Consciousness and self-regulation (Vol. 4; pp. 1-18). *New York: Plenum Press*.

Paolo, A. M., Tröster, A. I., Blackwell, K. T., Koller, W. C., & Axelrod, B. N. (1996). Utility of a Wisconsin Card Sorting Test short form in persons with Alzheimer's and Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology*, *18*(6), 892-897.

Siegelbaum, S. A., & Hudspeth, A. J. (2000). *Principles of neural science* (Vol. 4, pp. 1227-1246). E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.). New York: McGraw-hill.

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.