Study of habit learning impairments in Tourette syndrome and obsessive-compulsive disorder using reinforcement learning models

Vasco Conceição
Instituto Superior Técnico
Instituto de Medicina Molecular
Faculdade de Medicina da Universidade de Lisboa

Abstract

Psychiatric and neurological disorders are currently classified according to their symptomatic manifestations and not their intrinsic pathophysiology [89]. This leads to severe difficulties in patients' treatment and even diagnosis, and implies the need of developing new tools capable of objectively assessing the underlying neurophysiological impairments of such disorders [43] [80].

With this in mind, a reinforcement learning (RL) framework for the study of habit learning impairments during probabilistic classification learning task solving was created. This allowed the objective analysis of the behavioral data collected during the solving of the Weather Prediction Task (WPT) from several patients with Tourette syndrome (TS) and obsessive-compulsive disorder (OCD), and normal control (NC) subjects.

Two distinct datasets were analyzed, one regarding NC and OCD adults, and one concerning NC and TS children. In the first dataset, multiple adults had solved versions of the WPT with different input sequences, whereas all the analyzed children had solved a task with the same input and feedback sequence.

Several Q-learning models were created from the developed framework and fitted to the empirical data. These were compared through random effects Bayesian model selection (BMS) and, as a consequence, two models which reasonably explained the empirical data of the two datasets were selected.

Parametric comparison within the selected models across the NC and OCD adults and the NC and TS children was then performed. A one-tailed t -test revealed a tendency for increased switching in TS children ($p < 0.025$), which was also supported by a one-tailed Wilcoxon rank sum test ($p < 0.05$). RL-independent analysis of the data corroborated these results.

Key words

Dopamine; Obsessive-compulsive disorder; Reinforcement learning; Switching; Tourette syndrome

1. Introduction

"Memory is neither a unitary process nor does it serve a single function" [33].

In the past decades, multiple dichotomies in memory have been found, but the most significant contribution for this knowledge was probably obtained through the study of patient H.M., whose bilateral removal of the hippocampus led to anterograde and partial retrograde amnesia, without affecting his immediate memory span, motor skills, general intelligence, perceptual learning or even personality [49] [72].

Subsequent studies in animals and patients with amnesia corroborated this differentiation. These provided further evidence for an association between the structures in the medial temporal lobe (MTL), like the hippocampus, and explicit, or conscious, memory, and the existence of different forms of implicit learning [25] [30] [65] [75].

It is currently known, however, that grouping of memory systems in explicit and implicit is not formally correct. Unconscious engagement of episodic memory has been shown, and there is some evidence that factors like the temporal scale and cognitive complexity of learning, as well as the nature of mental representations, all characterize different types of memory [33]. Structures like the hippocampus, which were originally associated to the conscious engagement of episodic memory, have also recently been associated to the storing of short-term information and unconscious operation of the working memory (WM) [35], a volatile type of memory associated to the PFC which permits the online interpretation of the environment [1] [2] [19] [26] [27] [54].

Considering these findings, conditioning might be defined as an implicit memory formed through slow encoding of rigid associations [33]. Conditioning is divided into classical and instrumental. In the former, the outcome is independent of the actions performed by the subject and, in the latter, the outcome is determined by subjects' behavior. It is in this exact context that habit learning is defined [42].

Habit learning is a form of instrumental conditioning described by the gradual learning of stimulus-response (S-R) associations [35]. This is done outside conscious awareness, and typically results in rigid behaviors which are insensitive to the manipulations of the values of the outcome [21] [4].

In addition to habit learning, instrumental learning comprises the learning of stimulus-action-outcome (S-A-O) associations. These, conversely to S-R associations, are always sensitive to the values of the outcome [42] [78] [82].

The learning of both these associations has been intrinsically associated to the basal ganglia (BG), which are divided in the striatum (caudate nucleus, putamen and nucleus accumbens), and globus pallidus (globus pallidus interna and externa) [42]. In particular, the learning of S-R associations has been associated to the dorsolateral striatum circuitry and the learning of S-A-O associations to the circuitry involving the dorsomedial striatum and the pre-frontal cortex (PFC) [17] [44].

The BG receive several projections from the cortex and are involved in multiple loops, most notably grouped in the motor, associative and limbic circuits [43]. Additionally, they receive projections from the dopaminergic nuclei in midbrain: the ventral tegmental area (VTA) - which also projects densely to the frontal cortex and to the hippocampus -, and the substantia nigra pars compacta (SNr) [42].

These dopaminergic neurons fire in response to prediction errors (PE) - the difference between the new and the previously expected sum of future reinforcements -, providing the necessary framework for reward learning to occur [50] [71]. The key role of the BG in learning, as well as in more general processes of action selection is, precisely, mediated by dopamine [11] [27] [35] [43] [48] [54] [55] [69].

The role of the BG in reward learning and action selection

Strong evidence suggests that positive PEs are quantitatively represented by the magnitude of the phasic bursts of the dopaminergic neurons [25] [44], but the mechanisms underlying the encoding of negative PEs are not that well understood [44]. Dopaminergic neurons present a low tonic firing, and so,

encoding of the negative PEs by the magnitude of the phasic bursts does not seem to very plausible, since fluctuations below the baseline level would necessarily have limited power [43]. Some evidence exists that the negative PEs might be coded by the duration of the pause in the firing of dopaminergic neurons [3]. Still, more evidence is needed, as hypotheses implying a role for serotonin and the contribution of different dopaminergic populations to the encoding of negative PEs have also been postulated [44].

Irrespective of the exact mechanisms which are coding negative PEs, it is known that learning can only occur due to synaptic plasticity [6] [23] [42], a process to which dopamine has been strongly associated throughout the brain [34]. Hereafter, it is no surprise that dopaminergic input is crucial to reward learning, and that the BG play a central role in processes of action selection.

Whether the BG generate actions or select those being represented at the cortex is still uncertain [43] but, regardless of that, three distinct pathways involved in action selection have been identified between the cortex and the BG output structures: the direct, indirect and hyper-direct pathways (cf. figure 1) [43] [82]. These pathways have been commonly associated to the motor and associative loops. Nevertheless, there is also some evidence concerning their implication in the limbic loop [43]. Due to the BG projecting into the cortex through the thalamus, these loops are also commonly designated as the cortico-basal ganglia-thalamo-cortical (CBGTC) loops [43].

The direct (or Go) and indirect (or NoGo) pathways are known to mediate action selection through selective facilitation or inhibition of specific actions, respectively [82]. The hyper direct pathway, on the other hand, is hypothesized to mediate the action selection process through generalized inhibition of all actions. Thus, it is also commonly designated as the Global NoGo pathway [43].

Striatal neurons in the direct pathway express mostly dopamine D1 receptors, while striatal neurons in the indirect pathway predominantly express D2 receptors [43]. The D1 receptor presents lower affinity to dopamine than the D2 receptor and, for this reason, physiological tonic levels of dopamine lead to a higher binding of dopamine to the D2 receptor. Contrarily to the D1 receptors, which are mostly excitatory, D2 receptors are inhibitory [43]. Therefore, in the absence of PEs and homeostatic conditions, both pathways are tonically inhibited.

The direct pathway involves the cortex, striatum, thalamus, globus pallidus interna (GPi) and SNr, and it promotes the selection of the adequate action(s) for a given context through thalamus disinhibition [43]. When a PE occurs, the increased dopaminergic release in the striatum leads to a higher binding of dopamine to the D1 receptor. As a consequence, the Go neurons in the striatum inhibit the GPi and SNr and, since the projections from the GPi and SNr to the thalamus are inhibitory, the thalamus will have an increase in activation. Due to the excitatory feedback loop between the cortex and thalamus, Go learning and/or action facilitation will consequently occur. During this process, the indirect pathway is inactivated due to the binding between the dopamine and the striatal D2 receptors.

In the indirect pathway, the striatum projections terminate in the globus pallidus externa (GPe), and the GPe projections terminate in the GPi and SNr. Since both of these are inhibitory, when a negative PE occurs, the dopamine dip causes GPe inhibiting and, consequently, an increased inhibition of the thalamus (cf. figure 1). Hence, this pathway promotes NoGo learning and/or action inhibition. For the motives explained above, the Go pathway is not activated during this process.

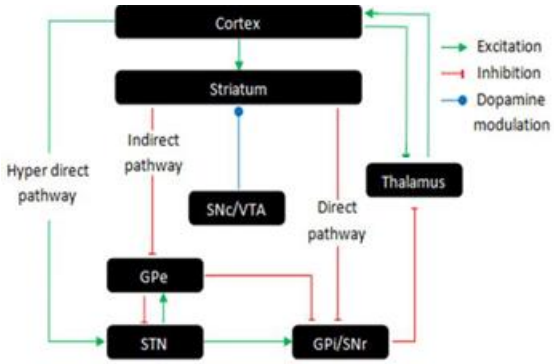

Figure 1: Simplified scheme of the anatomy of the CBGTC loops (adapted from [82]). [SNc: substantia nigra pars compacta; VTA: ventral tegmental area; GPe: globus pallidus externa; STN: subthalamic nucleus; GPi: globus pallidus interna; SNr: substantia nigra pars reticulata].

Consistently with these processes, which are triggered by variations in the phasic levels of dopamine, higher levels of tonic dopamine are associated to a Go bias, and lower tonic levels to a NoGo bias [43].

These mechanisms are intrinsically associated to the neurophysiology of habit learning [25] [44], but also to the gating mechanism underlying WM functioning, among others [27] [54]. However, other connections between dopamine binding in the striatum and action selection processes have also been established [11]. Striatal hyperdopaminergia, in particular, has been associated to an increased switching between competing responses, independently from reward history (possibly through increased binding to the D2 receptors in the caudate body) [11].

Considering this information, it is straightforward to understand the importance of this circuitry and that, for these mechanisms to be physiologically functioning, an intact dopaminergic system is essential.

While Parkinson is probably the most obvious example on how this circuitry is impaired, due to its well-documented associated loss of dopaminergic neurons [25] [56] [69], numerous neurodevelopmental disorders have been also reported to have this circuit severely compromised.

The pathophysiology of TS and OCD

TS and OCD are neurodevelopmental disorders which, like attention deficit hyperactivity disorder (ADHD), among others, affect the frontostriatal system [8]. They present a substantial genetic component and, not surprisingly, are highly comorbid [46] [52]. Environment and neurochemical deregulation are thought to play a key role on the onset of these conditions [8], but their etiology is still far from fully understood [8] [9] [45] [47] [57].

Patients suffering from these conditions typically have intact declarative memory, but subpar performances in most of the tasks which require a rapid and flexible adaptation of behavior to cope with changing contingencies [8]. In addition to severe deficiencies in motor control, deficits in attention, switching, inhibition and selection have all been identified in these patients [8]. The severity of the symptoms exhibited by these patients is, nonetheless, very variable, and a continuum of nocuous behaviors is thought to exist within each disorder, and even between these and other disorders [7] [10] [45] [46].

Patients suffering from TS usually manifest distinctive motor and vocal tics, but several other symptoms, as well as

comorbidities, have been associated to these patients. ADHD and OCD, in particular, may be present in up to 50% of the patients diagnosed with TS [46], if not more [7] [10], and TS symptoms may be motor, behavioral and cognitive [9]; hence the characterization of this disorder as highly heterogeneous.

Decreased inhibitory output from the BG has been associated with the pathophysiology of TS, and excessive activation of the direct pathway has also been hypothesized to be in the origin of the impaired inhibition of tics [8] [10] [43] [45]. These, coupled with the increased density of the dopamine transporter in TS patients, suggest that TS is intrinsically associated to a hyperinnervation of the striatum and/or an increased sensitivity of dopamine receptors [52]. The aggravation of symptoms following the exposure of TS patients to agents like L-Dopa, which increase dopaminergic activity, also supports these conclusions [9] [40].

Of the TS patients, some of the most problematic cases are those who suffer from comorbid ADHD [9]. These subjects tend to present, on average, more severe symptoms, and require very distinct treatments, depending on whether the TS or the ADHD symptoms are the most pronounced [8] [9] [10] [40] [47].

Like TS, OCD is extremely heterogeneous. The main characteristics of this disorder are, as the name might suggest, the manifestation of repetitive thoughts, or obsessions, and repetitive behaviors, or compulsions, by the patients [58] [80]. Deficiencies in the dopaminergic, serotonergic and glutamatergic systems are all known to contribute for the pathophysiology of this disorder, but how exactly these systems are impaired in each of the OCD subtypes is unknown [58]. In addition to the dissimilarity between childhood and adult-onset OCD, several other distinctions within this disorder have been made. It is currently believed that different clusters of symptoms within OCD exist and that, depending on which cluster is exhibited by a particular subject, different neural circuits may be compromised [58].

The standard pharmacological treatment for this condition involves selective serotonin-reuptake inhibitors (SSRIs), but dopamine antagonists, for instance, are also commonly used [10] [57] [80]. In addition, ADHD and mood disorders are often comorbid to OCD. This heterogeneity has been corroborated by multiple neuroimaging studies [58].

There is, therefore, an obvious need of better characterizing these disorders; a need which has been recently addressed through novel computational approaches [43]. PEs of different reinforcement learning models have been correlated with the firing of dopaminergic neurons in the midbrain [42] [44] [50] [51] [68] [71], and several other physiological processes concerning learning and action selection have been successfully captured by these models [11] [15] [28] [29] [56] [69] [82]. While there is still a long way to go before a reliable pathophysiology-based system for the classification of mental disorders can be developed, these computational approaches constitute, therefore, an important and promising tool to fulfill this goal [16] [42] [43] [80].

2. Methods

Frontostriatal impairments in TS and OCD patients were studied through behavioral analysis of the data collected from these, alongside with control subjects, during the solving of the WPT, a probabilistic classification learning task.

The Weather Prediction Task

In the WPT, a set of fourteen inputs is probabilistically associated to two possible weather outcomes, rain or shine [25] [73]. Four distinct elementary cues are used in this task, and each of these is associated with a fixed probability to a given outcome. At each trial, combinations including up to three of the elementary cues are shown and, depending on the particular cues which are involved in each combination, rain or shine may be predicted more or less accurately by these. The objective of this task is, precisely, to learn how to predict the outcome most strongly associated to the fourteen different cue combinations, that is, the inputs [25] [45] [74].

Due to the probabilistic nature of the WPT, a subject may correctly identify the outcome which was most strongly associated to a given input but receive a negative feedback as a consequence. Consistently, incorrect responses may also be rewarded. Regardless of the received feedback, only responses which selected the output most strongly associated to a given input are considered correct, though. For these reasons, when the first version of this task was designed, declarative memory of recent events was not expected to contribute to an increased performance of the subjects [32] [38].

Assuming the validity of this prior assumption, healthy subjects should exhibit an increased accuracy throughout the task, as they gradually learn the S-R associations of interest by incorporating the corresponding reward feedback, whereas patients with BG dysfunctions should present sub-optimal performances. It is currently known that this is not exactly true, though. MTL structures have been shown to compete with structures involved in habit learning during this task [61] [62] [63] [64] [75], leading to the employment of strategies different than those which were originally expected [25] [32] [73] [74]. Additionally, as in any other behavioral task, the contribution from WM to WPT solving cannot be disregarded either [1] [12].

This task presents, therefore, multiple problems. Nevertheless, it was hypothesized that useful information regarding habit learning could be obtained from the analysis of the WPT data, since previous studies had been able to successfully establish a connection between the habit learning impairments in Parkinson and Tourette patients, and the WPT performance of those subjects [25] [37] [45] [73].

Neurobehavioral data

The analyzed WPT data was divided in two datasets: an adult dataset, and a children dataset. All subjects had solved a task with 90 trials and the same probabilistic structure (cf. figure 2).

Cue 1	Cue 2	Cue 3	Cue 4	G(n)	H(n)	P(G)
0	0	0	1	10	2	.83
0	0	1	0	4	3	.57
0	0	1	1	7	1	.88
0	1	0	0	3	4	.43
0	1	0	1	5	1	.83
0	1	1	0	2	2	.50
0	1	1	1	3	1	.75
1	0	0	0	3	10	.23
1	0	0	1	2	3	.40
1	0	1	0	1	5	.17
1	0	1	1	2	1	.67
1	1	0	0	1	7	.13
1	1	0	1	1	2	.33
1	1	1	0	1	3	.25

Figure 2: Probability structure of the WPT (adapted from [45]). [G(n): number of shine outcomes; H(n): number of rain outcomes; P(G): probability of the input yielding shine as an outcome].

The first dataset contained data from 53 adults, who had either been assigned to a control group, or diagnosed with OCD (cf. table 1). In addition to the primary diagnosis of these patients, no other clinical information was made available. Several adults from both groups had solved versions of the WPT with different input sequences, but none of these subjects was excluded from the behavioral analyses.

	Number (%)	Age, years	Sex, M:F
NC	26 (0.49)	29.43 (7.85)	13:13
OCD	27 (0.51)	30.85 (8.47)	15:12

Table 1: Demographic and clinical data from the adults (≥ 18 years) of the normal control (NC) and obsessive-compulsive disorder (OCD) groups. Standard deviations of the ages are indicated between parentheses. [M: male; F: female].

The second dataset contained data from 36 children. Of these children, seventeen belonged to a control group, and nineteen had been diagnosed with TS. Thirty out of the thirty-six subjects which composed this dataset had solved a task with the same exact input and feedback sequence. The remaining six, however, had solved task versions with distinct sequences. Since all of these belonged to the NC group, they were excluded from the analyses. Like in the adult dataset, only primary diagnosis was considered when performing the behavioral analyses (cf. table 2).

	Number (%)	Age, years	Sex, M:F
NC	11 (0.37)	15.10 (2.77)	4:7
TS	19 (0.63)	11.68 (2.45)	16:3

Table 2: Demographic and clinical data from the children (<18 years) of the normal control (NC) and Tourette syndrome (TS) groups. Standard deviations of the ages are indicated between parentheses. [M: male; F: female].

Behavioral data concerning the different subjects was received in separate .edat files. These contained, amongst other variables, information regarding the inputs which were shown at each trial, subject's choices, and the reinforcements which they consequently received. All these files were pre-processed in E-Prime 1.2. Unanswered trials from both datasets were disregarded from the analyses.

A priori hypotheses testing

It was desired to study the performance of the NC, TS and OCD groups, to evaluate if there was evidence for impaired learning in the patients suffering from these disorders.

Direct interpretation of the behavioral data would only provide information regarding the evolution of the accuracy throughout the task, and the switching/perseverance trade-off between the shine and rain responses. Thus, to perform a complete study of the habit learning impairments in TS and OCD patients, besides executing this standard analysis, a computational model was developed. This was done in MATLAB, and provided the necessary framework for the creation of the different reinforcement learning submodels which were used in the analysis of the behavioral data.

Reinforcement learning

In a formal representation of RL, at a given instant, an agent is in a specific state, and it may select an action from the set of possible actions for that state. According to the selected action, the agent may then transition to a new state or not, as well as being reinforced or not. Due to the independence of this environment from the past, RL mechanics can be written in terms of an MDP [44].

Therefore, two possibilities exist for an agent to find the optimal policy, that is, the set of actions for each state which maximize his expected sum of future reinforcements. Either the agent tries to learn the MDP associated to the task that he is performing (which is unrealistic for tasks of higher complexity and without causal relation between the sequence of events), or the agent tries to learn the optimal policy regardless of the MDP [44]. These are designated as model-based and model-free approaches, and are linked to the learning of S-A-O and S-R associations, respectively [17] [44].

Following this reasoning, to fit the WPT behavioral data, the model-free approach was applied, and the different submodels were created under a framework similar to that of Q-learning models. This is a subclass of temporal-difference RL algorithms which estimates state-action values [77]. The state-action values are the expected values of performing a given action for a specific state. Biologically, these are hypothesized to be coded through dopamine-mediated cortico-striatal synaptic plasticity. Some recent imaging studies have supported the validity of this approach [51] [68].

The implementation of the model was, however, slightly more complex, so that multiple physiological processes could be tested and the necessary level of generalization for the analysis of the WPT was attained. The developed model does not use, for instance, the concept of state, as this is not straightforward during WPT solving. Instead, the model assumes the existence of several bits which, according to the inputs that the subjects receive, might be activated in a certain trial or not. Different learning hypotheses were also implemented.

For this reason, the computational model requires a set of options, built in a parameter-independent manner, to be specified by the user. The options of interest for this study are indicated in table 3. Options 1 and 2 were created to study the bit-response associations which had been internally represented and activated by the different subjects in reaction to the different inputs, and option 3 aimed at answering questions related to the context-dependent update of the different bits. Option 4, on the other hand, addressed the issue of choice autocorrelation; that is, the degree to which past choices had influenced subsequent responses, independently of how they were reinforced [16] [39] [70]. The first type aimed at quantifying the tendency to alternate between shine and rain responses, for the different inputs. The second type reflected a more basic instinct, which was hypothesized to be correlated with the dopamine levels on the striatum [11] [24] [66] [69].

Different specifications of these options, as well as of the parameters in use, defined the different submodels which were used in the analysis. For each subset of parameters, five submodels were created (cf. table 4).

The submodels were fitted to the behavioral data of each subject either through maximum likelihood or maximum *a posteriori* estimation. These two processes were accomplished through the careful use of the *fmincon* routine implemented in MATLAB, and allowed the estimation of the optimal parameters of each subject ($\hat{\theta}_M$, cf. equation 1).

$$\hat{\theta}_M = \arg\max_{\theta_M} P(D|M, \theta_M) \qquad (1)$$

Different submodels presented distinct bit activation patterns (cf. table 5) and sets of parameters (cf. table 6), but they were all submitted to an analogous fitting procedure (where null values were attributed to the absent parameters).

To reduce the risk of the calculated likelihoods underflowing the minimum value represented by a computer, the parameter estimation was done through log-likelihood (LLH) calculation. Hence, before the trial-by-trial fitting of each submodel was performed, the corresponding LLH and bit-response values $\big(Q(b,a)\big)$ were set to 0.

Positive and negative feedbacks were modeled as rewards (r_t) of values $+1$ and -1, respectively. Trial-by-trial operations were divided in two phases: action selection and internal values' update.

Option	Type 1	Type 2
I. Represented bits	All	Only elementary cues
II. Activated bits	All associated to a given input	Only fully conjunctive bits
III. Learning weights	$isactivated(input_{trial}, bit)$	$\frac{isactivated(input_{trial}, bit)}{\#active\ bits}$
IV. Choice autocorrelation	Previous answer for the same input	Previous answer

Table 3: User-dependent configurations of the computational model. Different inputs can be shown during the task, and so, at a given trial, a certain bit might be activated or not. Throughout the task, the different bits influence the action selection process according to their combination weights, and are updated according to their learning weights. The activation of a given bit following the exhibition of a certain input is determined by the corresponding $isactivated(input_{trial}, bit)$ boolean variable. Depending on the selected choice autocorrelation type, the last action done for the same input, or for the previous trial, might also influence the online decision process in reward-independent manner.

Option / Model	Represented Bits	Activated Bits	Learning Weights
1	All	All	Non-normalized
2	All	All	Normalized
3	All	Fully Conjunctive	-
4	Elementary	All	Non-normalized
5	Elementary	All	Normalized

Table 4: Characteristics of the submodels created for each subset of parameters. For a given parametric specification, these five types of models constitute a family.

Input \ Activated bits	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110
1: {0,0,0,1}														
2: {0,0,1,0}														
3: {0,0,1,1}														
4: {0,1,0,0}														
5: {0,1,0,1}														
6: {0,1,1,0}														
7: {0,1,1,1}														
8: {1,0,0,0}														
9: {1,0,0,1}														
10: {1,0,1,0}														
11: {1,0,1,1}														
12: {1,1,0,0}														
13: {1,1,0,1}														
14: {1,1,1,0}														

Table 5: Activation pattern of the different bits during WPT solving, for the different submodels. [grey cells: submodels which only represent and activate the elementary bits; colored cells: submodels which represent and activate all bits; diagonal cells: submodels which assume a fully conjunctive representation of the inputs].

In the action selection phase, firstly the combination weights ($w_{comb_{eff_t}}(i_t, bit_j)$) of the different bits are determined (cf. equation 2). This is done according to the input (i_t) which was shown at the corresponding trial (t), and so that the total value of the input-action association is always between the maximum and minimum rewards received. Then, the values of the preferences for both actions are calculated (cf. equation 3).

$$w_{comb_{eff_t}}(i_t, b) = \frac{isactivated(i_t, b)}{\#active\ bits} \quad (2)$$

$$p(i_t, a) = \sum_j Q_{t-1}(b_j, a).w_{comb_{eff_t}}(i_t, b_j) \quad (3)$$

Finally, the probabilities of performing the different actions are calculated. This was implemented in the form of a *softmax* equation [16] [77], cf. equation 4.

$$\forall_a \quad Prob(a) = \frac{e^{p(i_t,a).\beta_{eff_t}(a)+k.f(a)}}{\sum_j e^{p(i_t,a_j).\beta_{eff_t}(a_j)+k.f(a_j)}} \quad (4)$$

In the above equation, k is the choice autocorrelation parameter, $f(a_l)$ are boolean variables indicating whether k should be added to the preferences for the respective actions or not (in agreement with table 3), and β_{eff_t} are functions which allowed that positive and negative preferences could be weighted through different parameters. The latter were also defined as a function of t since the exploration-exploitation strategies of the subjects could be dynamic. This increased flexibility was modeled through the use of four parameters: the inverse temperatures ($\beta_G, \beta_L,\ \beta$), and the exploration-exploitation modulator (ϵ), cf. equations 5 and 6.

$$\beta_{eff_0}(a) = q_1(a).\beta_G + q_2(a).\beta_L + q_3(a).\beta \quad (5)$$

$$\beta_{eff_t}(a) = \beta_{eff_0}(a) + \epsilon.(t-1) \quad (6)$$

Where q_l are boolean variables determined by the submodel in use (cf. equations 7 to 9).

$$q_1(a) = \begin{cases} 1, & if\ p(a) > 0 \\ 0, & otherwise \end{cases} \quad (7)$$

$$q_2(a) = \begin{cases} 1, & if\ p(a) < 0 \\ 0, & otherwise \end{cases} \quad (8)$$

$$q_3(a) = \begin{cases} 1, & if\ (\beta_G = 0 \wedge \beta_L = 0) \\ 0, & otherwise \end{cases} \quad (9)$$

The implementation of different inverse temperatures was mainly motivated by two articles which summarized the importance of considering the differential contributions of dopamine in incentive and choice, besides learning [13] [43].

After probability calculation, and in order to fit the behavioral data, the model selects the action which was performed by the subject at that trial. Subsequently, the log-likelihood (LLH) is increased by the logarithm of that probability (cf. equation 10).

$$LLH_t = LLH_{t-1} + \log\big(Prob(a_t)\big) \quad (10)$$

No.	Parameter	**Identification.** Description / usage	Range
1	α_G	**Positive striatal learning rate.** This parameter models reward learning occurring through the direct pathway of the CBGTC loops. It is associated to positive prediction errors.	[0,1]
2	α_L	**Negative striatal learning rate.** This parameter models reward learning occurring through the indirect pathway of the CBGTC loops. It is associated to negative prediction errors.	[0,1]
3	α	**Single striatal learning rate.** This parameter assumes that the same mechanism is used for the incorporation of positive and negative feedback. Therefore, its use is incompatible with the use of α_G and α_L. These alternative hypotheses should be evaluated through model comparison procedures.	[0,1]
4	ϕ	**Striatal learning rates' discount factor.** This parameter accounts for the nonstationary of the learning rates. It models their hypothesized decrease with time.	[0,0.5]
5	β	**Single inverse temperature.** This parameter weights the preferences for all possible actions during the action selection process.	[0,20]
6	β_G	**Positive inverse temperature.** This parameter assumes that different mechanisms are used in the evaluation of the positive and negative preferences for the different actions. For this reason, it is incompatible with β. It weights the positive preferences.	[0,20]
7	β_L	**Negative inverse temperature.** This parameter is similar to the positive inverse temperature, with the exception that it weights the negative preferences.	[0,20]
8	ϵ	**Exploration-exploitation modulator.** This parameter assumes that as the task is solved, the action selection process is becoming increasingly more deterministic. It is used in the update of the values of the inverse temperatures.	[0,0.2]
9	k	**Choice autocorrelation parameter.** This parameter determines the degree to which a past choice is influencing future action selections, independently of the outcome from that choice. Positive values favor perseverance and negative values switching.	[-20,20]

Table 6: Brief description of the parameters used during the RL-based analysis of the WPT behavioral data. The ranges of the different parameters (right column) were chosen to guarantee that all parameters had physiological meaning. Optimal values outside these boundaries would indicate a lack of quality of the models to correctly fit the behavioral data.

The internal reward (int_{rew_t}) is then obtained. This was assumed to be equal to the respective reward, for all subjects.

After this step, the internal reward is used to calculate the PEs associated to the activated bit-response values. Since the sequence of future inputs is independent from the choice of the subject at each trial, the PEs only depend on the immediate rewards and the previously expected values (cf. equation 11). Like in the action selection phase, the bits of interest are defined by the input shown at that trial (cf. table 3). However, only the associations involving the selected action may be activated.

$$\delta_{str_t}(bit, a_t) = int_{rew_t} - Q_{t-1}(bit, a_t) \quad (11)$$

Next, the effective striatal learning rates for each of the bit-action pairs of interest are calculated (cf. equations 12 and 13). The chosen implementation addressed the existence of a direct and an indirect pathway in the CBGTC loops through the modeling of a positive learning rate (α_G, associated to the direct pathway), and a negative learning rate (α_L, associated to the indirect pathway). A single learning rate (α) was also modeled, in order to test the validity of these assumptions. Additionally, a parameter ϕ that accounted for parametric nonstationarity was included in the model. This was done because different studies had reported that the learning rates were a function of the uncertainty regarding the stimuli [18] [36], and implemented under a simplified scheme, similar to the one used in [53].

$$\alpha_{str_0}(b, a_t) = q_4(b, a_t).\alpha_G + q_5(b, a_t).\alpha_L + q_6(b, a_t).\alpha \quad (12)$$

$$\alpha_{str_t}(b, a_t) = \alpha_{str_0}(b, a_t).\exp\big(-\phi.(t-1)\big) \quad (13)$$

The boolean variables in equation 12 are defined according to equations 14 to 16.

$$q_4(b, a_t) = \begin{cases} 1, & if\ \delta_{str_t}(b, a_t) > 0 \\ 0, & otherwise \end{cases} \quad (14)$$

$$q_5(b, a_t) = \begin{cases} 1, & if\ \delta_{str_t}(b, a_t) < 0 \\ 0, & otherwise \end{cases} \quad (15)$$

$$q_6(b, a_t) = \begin{cases} 1, & if\ (\alpha_G = 0 \wedge \alpha_L = 0) \\ 0, & otherwise \end{cases} \quad (16)$$

Finally, the Q-values of interest are updated (cf. equation 17).

$$Q_t(b, a_t) = Q_{t-1}(b, a_t) + \cdots \\ \ldots \alpha_{str_t}(b, a_t).w_{learn_{eff_t}}(i_t, b).\delta_{str_t}(b, a_t) \quad (17)$$

With the created models, it was desired to evaluate the quality of the implemented learning approaches (through the use of different model types, cf. table 4), but also the relevance of the different parameters, for the NC, TS and OCD groups. It was hypothesized that better task performances could be associated to parametric nonstationarity, and that, due to the striatal hyperdopaminergia and dopamine hypersensitivity hypotheses of TS, TS patients could present higher activation of the direct pathway than the indirect one and/or increased switching between shine and rain responses, when compared to control subjects. This evaluation was done through the application of a sequence of model fitting and model comparison procedures to each dataset.

To perform an unbiased model comparison across groups, the model evidence (ME) ($P(D|M)$, cf. equation 18) should be used [5] [16] [76].

However, due to the analytical intractability of the integral required for its calculation, the use of this exact value was not possible.

$$P(D|M) = \int P(D|M, \theta_M).P(\theta_M|M)\, d\theta_M \quad (18)$$

Several hypotheses existed to (partially) overcome this issue, and so estimate the model evidence, but all of them present severe cons. To perform this procedure, three approximations to the ME were used, the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the Laplace approximation (LA).

Both the AIC and the BIC present an accuracy term, which is given by the maximum likelihood estimate, and a second term which tries to correct for model complexity (cf. equations 19 and 20, where n and m identify the number of parameters of the model and the number of data points to be fitted, respectively).

$$AIC = \log\left(P\left(D\middle|M,\hat{\theta}_M\right)\right) - n \quad (19)$$

$$BIC = \log\left(P\left(D\middle|M,\hat{\theta}_M\right)\right) - \frac{n}{2}\log(m) \quad (20)$$

The main advantage of using those approximations was to circumvent the use of the priors. Some others existed [5] [20] [60], but both criteria were far from ideal. The complementarily use of these criteria provided, nonetheless, important information due to the under and over penalizing natures of the AIC and BIC, respectively [59] [60].

The LA presents a better correction for model complexity than the AIC and BIC. This is due to the covariance-dependent nature of its complexity penalty (cf. equation 21, where H is the hessian of the negative logarithm of the product between the likelihood of the data and the prior over the parameters). However, some strict assumptions are necessary for the use of this approximation, namely that the likelihood distribution around the optimal parameters ($\hat{\theta}_M$) is approximated by a multivariate Gaussian centered on that estimate [5] [16].

$$LA = \log\left(P\left(D\middle|M,\hat{\theta}_M\right)\right) + \log\left(P\left(\hat{\theta}_M\middle|M\right)\right) + \cdots$$
$$\ldots\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|H| \quad (21)$$

Due to the explicit dependence of the LA on the priors over the parameters, the $\hat{\theta}_M$ used in this approximation must be obtained through maximum *a posteriori* and not maximum likelihood estimation. During this analysis, a flat prior over the range of the parameters (cf. table 6) was assumed for each parameter.

The Gaussian assumption required by this approximation was not always reasonable during the executed analysis. This was due to the quantity of hypotheses being tested and to the existence of some slight problems in the models. This led to the obtainment of some nonsense values for this approximation.
The increased sensitivity of the LA to the covariance between the parameters was also problematic in some situations. The reason behind this is particularly evident when considering equation 22, a slightly modified version of equation 21.

$$P_{LA}(D|M) = P\left(D\middle|M,\hat{\theta}_M\right).P\left(\hat{\theta}_M\middle|M\right).((2\pi)^d|H^{-1}|)^{1/2} \quad (22)$$

In cases of parameter degeneracy, where multiple combinations of parameters produce the same likelihood values, the term $|H^{-1}|$ massively increases, causing an over-inflation of the model evidence [5]. The use of the LA in such cases would bias the model comparison procedure towards the selection of nonsense degenerate models.

The comparisons between the different models, for the adult and children datasets, were made through the use of a hierarchical model, which treated the models in study as random effects (cf. figure 3) [16] [76].

This model assumes that each subject (i) is characterized by a set of binary variables (m_{ik}) which indicate the model (k) that is assigned to that subject, and that those variables are generated by a multinomial distribution with parameters $\boldsymbol{r}$ (the model probabilities to be estimated). Moreover, it assumes that the latter are described by a Dirichlet distribution, with parameters $\boldsymbol{\alpha}$.

The hierarchical model only requires the (approximated) model evidences ($p(y_i|m_k)$) to be observed, and its inversion yields the values of $\boldsymbol{\alpha}$ which, when subtracted the prior, reflect the effective number of subjects for whom a given model generated the data [76].

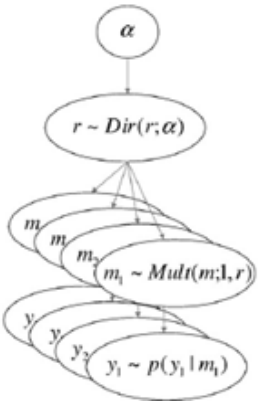

Figure 3: Random effects generative model for the data of the population (adapted from [76]).

Inversion of the hierarchical model allowed quantification of the belief of a given model being more likely than the remaining $K-1$ tested models, in the form of its exceedance probability (ϕ_k, cf. equation 23).

$$\forall j \in \{1 \ldots K | j \neq k\},\ \phi_k = P(r_k > r_j|y,\alpha) \quad (23)$$

The exceedance probabilities (EPs) of the different models sum to one [76], but they are not protected against the possibility of the observed differences in model frequencies having been generated by chance. This was corrected through calculation of the protected exceedance probabilities ($\tilde{\phi}_k$) [67], cf. equation 24 (where BOR is the Bayesian omnibus risk of having wrongly assumed that the differences in model frequencies were real, when they were simply due to chance [67]).

$$\tilde{\phi}_k = \phi_k.(1-BOR) + \frac{1}{K}.BOR \quad (24)$$

This hierarchical model also provided the tools for comparison between families of models, which was essential to evaluate different subsets of parameters, regardless of the model types. This was done by calculating the Dirichlet parameters corresponding to each family (α_{F_j}), and then using these family parameters to calculate the protected exceedance probabilities (PEPs) in a similar manner to the one described above [76]. The different α_{F_j} are obtained by summing the α_k of the models contained in each family. The values of the latter depend, however, on the Dirichlet prior. More specifically, when the models are divided in families, this prior is usually assumed to be flat on the families, and not on the models, which might lead to different results in terms of model selection [14]. This was indeed the case, during the performed analysis.

This methodology also allowed the quantification of the likelihood of distinct groups having different model frequencies ($P(H_{\neq}|y)$) [67].

To do so, two hypotheses were postulated: $H_{=}$, which assumed that all groups presented the same model frequencies, and $H_{\neq}$, which claimed the opposite. Being H_1 the assumption that there is indeed a real difference in model frequencies, $P(y|H_{=}) = P(y|H_1)$. According to $H_{\neq}$, on the other hand, the subsets (y_s) from the different groups are marginally independent. Following this reasoning, $P(H_{\neq}|y)$ was easily calculated from equations 25 to 27 [67].

$$P(y|H_{\neq}) = \prod_{s=1}^{S} P(y_s|H_1) \quad (25)$$

$$P(y|H_{=}) = P(\cup_{s=1}^{S} y_s|H_1) \quad (26)$$

$$P(H_{\neq}|y) = \frac{P(y|H_{\neq})}{P(y|H_{\neq})+P(y|H_{=})} \quad (27)$$

This between-group comparison had to be carefully applied, though. This was due to this approach being slightly biased towards the $H_{\neq}$ assumption because of the uncertainty regarding individual log-evidences and the within-group variability in terms of models [67].

All these Bayesian model selection (BMS) processes were done in MATLAB. To do so, the toolbox described in [14] was downloaded and adapted, so that the PEPs of the analyzed families and models were calculated, and the output from the *VBA_groupBMC* was more easily interpreted.

The exact sequence of the model comparison procedures was not defined *a priori,* so that an unbiased study could be performed. Instead, these procedures were begun by fitting and comparing models with one learning rate and a single inverse temperature, and they proceeded from there. This was the initial step for both datasets because it would not introduce any noise (due to a difference in the number of parameters among models) in the comparison when the AIC or BIC were used, and because it was expected that the LA would be relatively well supported for models with these subsets of parameters.

All results were mainly interpreted in terms of PEPs, as supported in the most recent articles regarding this methodology [76].

BMS procedures were, however, complemented through the execution of likelihood ratio tests (LRTs). These evaluated the significance of the within-subject increase in likelihood caused by the addition of n extra parameters to a given model, by testing the null hypothesis that the simpler model was the correct one [16].

Two distinct cases could occur during the RL-based analysis. In the first case, no learning strategy would stand out and differences between groups would be evaluated through comparison of the different families of models. Alternatively, a single model could be proven to explain the data better than all the other models and significantly better than what it was expected by chance. In this situation, assumptions regarding the hypotheses of interest would be evaluated through parametric comparison across groups [16] [67] [76].

3. Results

Standard behavioral analysis

To study the subjects' accuracy throughout the WPT, block analysis was performed, as suggested in [45].

In the first dataset, no trend for the evolution of the percentage of correct answers across the task was visible for the NC adults or the OCD adults. A mixed ANOVA did not provide conclusive results either. Both groups were found to have a performance significantly above chance for every block, but no differences in accuracy across blocks or groups were detected. However, a significant block by group interaction ($F(8,44) = 2.361, p < 0.05$) was detected. To confirm that this interaction had not confounded any meaningful result, repeated-measures ANOVAs were applied to the NC and OCD data separately, but these did not show any difference between blocks.

An analogous procedure for the children dataset did not detect any difference between the NC and TS groups either.

Regarding the switching/perseverance trade-off, a different approach was used. The number of shifts in action selection between consecutive trials for each subject (X_i) was assumed to follow a Binomial distribution with parameters n_i and p_i, respectively the number of valid trials and the probability of switching between rain and shine. According to the number of valid trials and the number of shifts between shine and rain responses done by each subject (k_i), Bayesian estimation was then performed to approximate the p_i's, cf. equation 28.

$$\widehat{p_i} = E[Beta(k_i + 1, n_i - k_i + 1)] \quad (28)$$

The distributions of the $\widehat{p_i}$'s obtained for the NC and OCD adults, and the NC and TS children were then subjected to a normality test, the Lilliefors test [41]. This was not significant for any of the populations, and so, the switching proportions found for the individuals were compared through t-tests. A one-tailed t-test revealed increased switching for the TS children when compared to the NC children ($p < 0.05$), but would also reveal increased switching in the OCD adults when compared to the NC adults ($p < 0.05$). A non-parametric one-tailed Wilcoxon rank sum confirmed the existence of increased switching in the TS children ($p < 0.05$).

RL-based analysis

BMS between the five models of each family with one parameter and a single inverse temperature – the (α, β), (α_L, β) and (α_G, β)-families - yielded a PEP over 95% for the (α_G, β)-family, irrespectively of the ME approximation used, for both datasets.

To evaluate which parameter, besides the positive learning rate and the single inverse temperature, was the most informative, comparisons between the 3-parameter models including α_G and β were subsequently done. To do so, models from the $(\alpha_G, \alpha_L, \beta)$, $(\alpha_G, \beta_G, \beta_L)$, $(\alpha_G, \epsilon, \beta)$, (α_G, k, β) and (α_G, ϕ, β)-families were fitted and compared. Since two possibilities of choice autocorrelation existed (cf. table 3), 6 families were compared.

Some of the models from these families did not provide a good fit to the data, which led to that their likelihood distributions did not fulfill the requirements for the use of the LA. This degraded the quality of this approximation significantly, and invalidated its use.

Comparison of the models from these families using the BIC indicated that the $(\alpha_G, \alpha_L, \beta)$-family and the second (α_G, k, β)-family were the ones which best explained the data for the adult and children datasets (PEPs> 0.95), respectively. The latter assumed choice autocorrelation between consecutive trials

BMS between these models was also separately performed for the NC and OCD adults in the first dataset and for the NC and TS children in the second. In the former, no significant difference between groups was detected, as the $(\alpha_G, \alpha_L, \beta)$-family was favored for the NC and OCD adults, consistently with what had been observed for the simultaneous analysis of the whole dataset. In the children dataset, however, BMS between these models revealed extremely different model frequency patterns for the NC and TS children. The $(\alpha_G, \alpha_L, \beta)$-family was favored for the NC group (EP> 0.95), and the second (α_G, k, β)-family for the TS one (PEP> 0.95).

In agreement with the above results, the (α, β), (α_L, β), (α_G, β) and $(\alpha_G, \alpha_L, \beta)$-families were then compared for the adult dataset. For the children dataset, these and the second (α_G, k, β)-families were compared. Conversely to the previous comparisons between the 2-parameter and 3-parameter models, the use of the AIC and BIC led to very different results in these.

Bayesian model comparisons (BMCs) between the twenty models from the (α, β), (α_L, β), (α_G, β) and $(\alpha_G, \alpha_L, \beta)$-families were not conclusive for the adult dataset. Depending on the used approximation to the ME, these favored either the second models (cf. table 4) from the (α_G, β) or the $(\alpha_G, \alpha_L, \beta)$-families. All

other models presented negligible EPs and PEPs irrespectively of the used approximation. Thus, these two models were selected. Through BMS it was not possible, however, to decide which of these was the most correct. Since these were nested models, LRTs for each subject were performed (cf. figure 4). The parametric distributions obtained for these models were also analyzed. None of these provided definite evidence towards any of the models; hereafter, the 3-parameter model was used to continue the model comparison process. BMS between this and its nested 4-parameter models did not support a fourth parameter.

This meant that for the adult dataset there was only uncertainty regarding which of the second models from the (α_G, β) and $(\alpha_G, \alpha_L, \beta)$-families was the most correct. Between-group comparison did not support any difference between the frequencies of these models across groups $(p(H_{\neq}|y) < 0.5)$. Thus, to avoid losing important information, the latter was selected. This implied, however, that there was probably overfitting for some of the subjects.

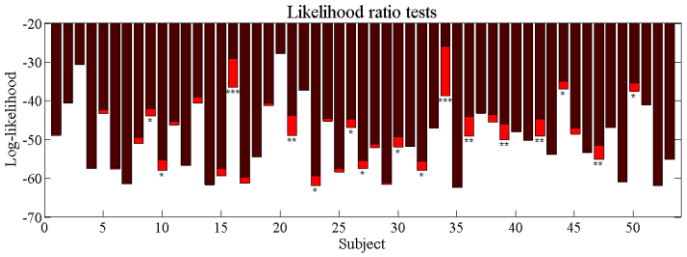

Figure 4: Log-likelihoods' distribution and results of the likelihood ratio tests applied to the second models of the (α_G, β) and $(\alpha_G, \alpha_L, \beta)$-families, for the adult dataset. The first 26 subjects belong to the NC group.

Regarding the children dataset, BMS between the (α, β), (α_L, β), (α_G, β), $(\alpha_G, \alpha_L, \beta)$ and the second (α_G, k, β)-families did not provide conclusive results either. BMS using the AIC selected the fifth model of the (α_G, k, β)-family for the full dataset and the TS children (PEP > 0.95), but BMS using the BIC yielded PEPs over 0.6 for the fifth model of the (α_G, β)-family for these same subsets. Moreover, BMS using the AIC and BIC for the NC group yielded PEPs of approximately $1/25$ for all models, with the AIC slightly favoring the 3-parameter families and the BIC slightly favoring the (α, β) and (α_G, β)-families (frequencies above chance). This suggested that modeling the negative learning rate and the choice autocorrelation parameter would be important for the NC and TS children, respectively. To evaluate if there was indeed a difference in model frequencies between the NC and TS groups, random effects between-group comparison was performed for the fifth models of the (α_G, β), $(\alpha_G, \alpha_L, \beta)$ and (α_G, k, β)- families. These were the only models which had presented non-negligible EPs for the full dataset, NC and TS subsets when the AIC was used.

Likelihoods of the NC and TS children having different model frequencies of 0.58 and 0.68 were obtained when using the AIC and BIC, respectively. These were considerable values but, due to the intrinsic bias of this measure towards the finding of a difference between groups, they were not considered to be enough to support a statement as strong as the existence of a difference in model frequencies between groups.

When the BMS using the BIC had been performed for the comparison between the six 3-parameter families, a PEP over 0.95 had been obtained for the (α_G, k, β)-family, and its fifth model. Hence, the model comparison procedure was continued by comparing the nested 4-parameter versions of this model. BMS using the BIC yielded a PEP over 0.95 for the $(\alpha_G, \alpha_L, k, \beta)$-model for the full dataset, and over 0.8 for the NC and TS subsets. Therefore, this model was confidently selected.

To confirm that no mistake had been made in only fitting and comparing the fifth models of these families, the remaining models from the $(\alpha_G, \alpha_L, k, \beta)$-family were fitted, and all models from this family were compared. A PEP over 0.95 was obtained for the fifth model, validating the previous steps.

Since no difference in model frequencies across groups had been proven, and there was still some uncertainty regarding the importance of modeling α_L and k, besides α_G and β, the fifth models from the (α_G, β) and $(\alpha_G, \alpha_L, k, \beta)$-families were compared. Once again, analysis of the likelihood and parametric distributions of these models were not conclusive. LRTs with a 95% confidence interval favored the 4-parameter model for 6 out of the 11 NC subjects and 8 out of the 19 TS subjects (cf. figure 5). This, coupled with the previous findings concerning the importance of modeling the negative learning rate and the choice autocorrelation parameter for the NC and TS children, sustained the selection of the fifth $(\alpha_G, \alpha_L, k, \beta)$-model for this dataset.

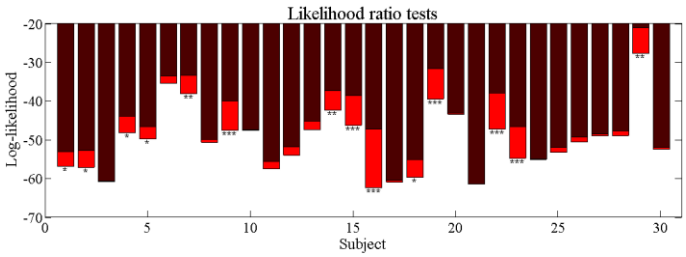

Figure 5: Log-likelihoods' distribution and results of the likelihood ratio tests applied to the fifth models of the (α_G, β) and $(\alpha_G, \alpha_L, k, \beta)$-families, for the children dataset. The first 11 subjects belong to the NC group.

To evaluate the quality of the models which had been selected for the two datasets, these were compared with the corresponding random models. BMS between the selected and the random models presented PEPs over 0.95 for the formers for both datasets, and irrespectively of the used approximation to the ME. Since the RL models had been created to capture the gradual learning of the S-R associations, and better performances in the WPT are usually associated to better learning of these associations, the connection between the goodness of fit of the models and the corresponding subjects' performance during the WPT was analyzed.

To do so, Pearson correlations between the average likelihoods per trial and the percentages of correct answers of each subject were performed (cf. figures 6a and 6b). Positive correlations between these were found for the adult and children datasets ($r = 0.84, p < 0.001$; $r = 0.76, p < 0.001$; respectively).

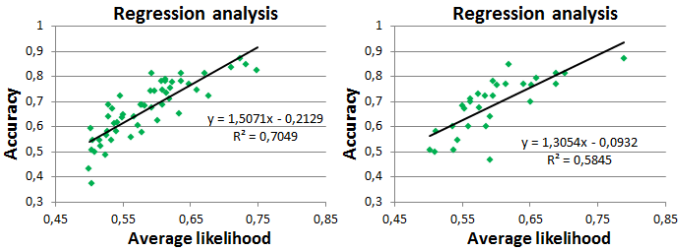

Figure 6: Regression of the average likelihoods into the mean accuracies of each subject, for the (a) adult and (b) children datasets.

The correlations between the mean accuracies and the mean probabilities of choosing the correct answer predicted by the models were then studied. Pearson correlations yielded significant positive correlations between these two quantities for the NC and OCD adults ($r = 0.57, p < 0.001$; $r = 0.64, p < 0.001$), but also for the NC and TS children ($r = 0.65, p < 0.001$; $r = 0.64, p < 0.001$). These quantities are depicted in figures 7 and 8.

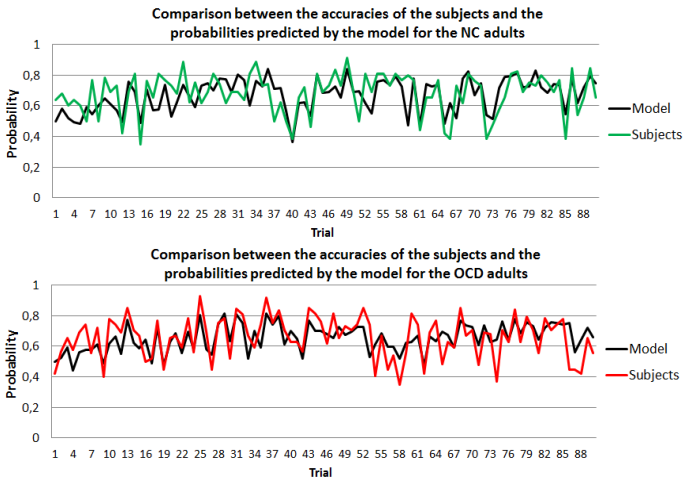

Figure 7: (a) Mean accuracies (green) and mean probabilities of performing the correct actions predicted by the model for the NC adults (black). (b) Mean accuracies (red) and mean probabilities of performing the correct actions predicted by the model for the TS children (black).

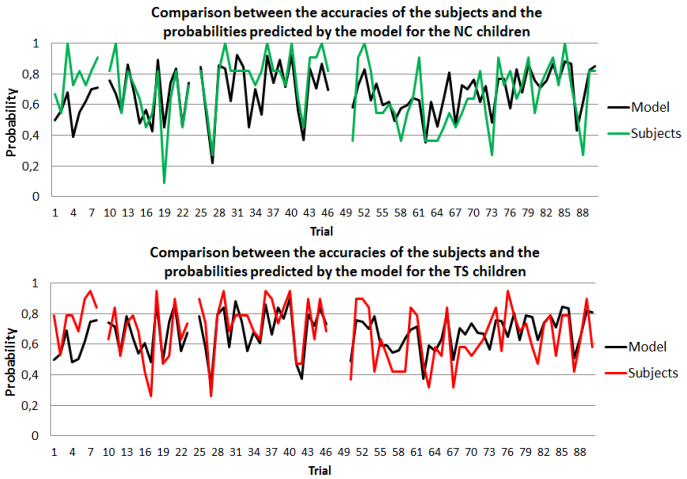

Figure 8: (a) Mean accuracies (green) and mean probabilities of performing the correct actions predicted by the model for the NC children (black). (b) Mean accuracies (red) and mean probabilities of performing the correct actions predicted by the model for the TS children (black). The 50/50 trials were omitted since there was no correct answer for these; hence, the discontinuities.

Subsequently, the selected models were used to perform parametric comparisons across groups. This was done through a summary statistics approach [16], but, before this was executed, the increased covariance between the learning rates and the inverse temperature was addressed [16] [69] [70]. As suggested in [69], new variables ($\alpha'_G = \beta.\alpha_G$, $\alpha'_L = \beta.\alpha_L$) were created from the optimal parameters,.

To test if the values of these two variables, or even the difference between them, were significantly different between the NC and OCD adults, and/or the NC and TS children, Wilcoxon rank sum tests were performed. These were done instead of the standard t -tests because the normality assumption for these variables had been rejected. No difference between groups was found. Additionally, Wilcoxon signed rank tests revealed that $\alpha'_G > \alpha'_L$ for all groups (p <0.05)

Finally, the switching/perseverance trade-off was analyzed. Lilliefors tests did not reject the hypothesis that k was normally distributed for the NC nor the TS children; hence, a one-tailed t-test was performed. This revealed a decreased value of k for the TS group (p <0.025); a finding which was corroborated by a one-tailed Wilcoxon rank sum test (p<0.05), and that confirmed the results obtained through RL-independent behavioral analysis.

4. Discussion

The fact that no evidence supporting the existence of gradual learning throughout the WPT or even a difference in performance across groups was found, precluded some of the *a priori* hypotheses to be tested. This might be, however, easily explained by the poor design of the task [62] [64]. Besides the CBGTC circuitry, several other regions have been shown to be activated during WPT solving; a fact which is highly associated to the existence of subpar performances [32] [73]. The lack of appropriate information concerning the medication status of the patients also complicated the analysis, since the adopted grouping of subjects might not have been the most correct.

The adopted BMS framework proved, nonetheless, to be very useful. This framework allowed all hypotheses to be tested without requiring prior assumptions on specific distributions of the models or the parameters on the populations to be made. The implementation of multiple learning strategies was also revealed to be a good approach, since the selected RL models assumed the existence of different strategies.

The found dissimilarity between the positive and negative learning rates supported the existence of different mechanisms for incorporation of positive and negative reward feedbacks. The values obtained for the negative learning rates require, however, careful analysis, as there were much more trials rewarded with positive than with negative feedbacks, and this might have led to a problem in the estimation of their exact values.

Considering these factors and the existence of multiple comorbidities, at least in the TS patients, it is not difficult to understand why no difference regarding the activation of the direct and indirect pathways was found across groups. Moreover, this suggests that some of the *a priori* hypotheses might have been over ambitious.

The existence of strong correlation between the goodness of fit of the selected models and the subjects' performances, and between the subjects' responses and models' predictions indicates, nonetheless, that the used methodology had considerable success in the simulation of subjects' behavior.

RL-based analysis of the switching/perseverance trade-off was also very satisfactory. In agreement with the direct behavioral analysis of the data, and consistently with our hypothesis, this analysis supported the existence of increased switching in TS children.

Striatal hyperdopaminergia had been previously associated with increased switching in marmoset monkeys [11] and Parkinson patients, who present a loss of dopaminergic neurons in the midbrain, had been shown to exhibit increased perseverance during behavioral task solving [69]; but, to my knowledge, this had not been reported in TS patients yet.

Inclusion of information regarding the patients' medication, comorbidities and symptom severity in the analyses should be addressed in the future, though. This would allow for a more accurate study of this and other hypotheses following an analogous methodology, but also the accurate use of hierarchical models of parameter estimation. The latter, in particular, would be very beneficial since these models explicitly deal with the within-subject variability on the parameter estimates.

Concerning model comparison, the use of better approximations to the ME (possibly calculated through Markov Chain Monte Carlo sampling [22] or variational Bayes techniques [31] [59] [67]) should also be targeted, so that all the model selection procedures can be formally validated.

Acknowledgements

I would like to thank all my friends and family for all the support and everything else. It has truly been a pleasure to study Biomedical Engineering at IST.

References

[1] Ashby, F. Gregory, and Jeffrey B. O'Brien. "Category learning and multiple memory systems." *Trends in cognitive sciences* 9.2 (2005): 83-89.

[2] Baddeley, Alan. "Working memory." *Science* 255.5044 (1992): 556-559.

[3] Bayer, Hannah M., Brian Lau, and Paul W. Glimcher. "Statistics of midbrain dopamine neuron spike trains in the awake primate." *Journal of Neurophysiology* 98.3 (2007): 1428-1439.

[4] Bayley, Peter J., Jennifer C. Frascino, and Larry R. Squire. "Robust habit learning in the absence of awareness and independent of the medial temporal lobe." *Nature* 436.7050 (2005): 550-553.

[5] Beal, Matthew James. *Variational algorithms for approximate Bayesian inference*. Diss. University of London, 2003.

[6] Bliss, Tim VP, and Graham L. Collingridge. "A synaptic model of memory: long-term potentiation in the hippocampus." Nature 361.6407 (1993): 31-39.

[7] Bloch, Michael H., et al. "Adulthood outcome of tic and obsessive-compulsive symptom severity in children with Tourette syndrome." *Archives of pediatrics & adolescent medicine* 160.1 (2006): 65-69.

[8] Bradshaw, John L., and Dianne M. Sheppard. "The neurodevelopmental frontostriatal disorders: evolutionary adaptiveness and anomalous lateralization." *Brain and language* 73.2 (2000): 297-320.

[9] Cavanna, Andrea E., and Hugh Rickards. "The psychopathological spectrum of Gilles de la Tourette syndrome." *Neuroscience & Biobehavioral Reviews* 37.6 (2013): 1008-1015.

[10] Cavanna, Andrea E., and Cristiano Termine. "Tourette syndrome."*Neurodegenerative Diseases*. Springer US, 2012. 375-383.

[11] Clarke, H. F., et al. "Orbitofrontal Dopamine Depletion Upregulates Caudate Dopamine and Alters Behavior via Changes in Reinforcement Sensitivity." *The Journal of Neuroscience* 34.22 (2014): 7663-7676.

[12] Collins, Anne GE, and Michael J. Frank. "How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis." *European Journal of Neuroscience* 35.7 (2012): 1024-1035.

[13] Collins, Anne GE, and Michael J. Frank. "Opponent Actor Learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive."

[14] Daunizeau, Jean, Vincent Adam, and Lionel Rigoux. "VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data." *PLoS computational biology* 10.1 (2014): e1003441.

[15] Daw, Nathaniel D., et al. "Model-based influences on humans' choices and striatal prediction errors." *Neuron* 69.6 (2011): 1204-1215.

[16] Daw, Nathaniel D. "Trial-by-trial data analysis using computational models."*Decision making, affect, and learning: Attention and performance XXIII* 23 (2011): 3-38.

[17] Daw, Nathaniel D., Yael Niv, and Peter Dayan. "Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control."*Nature neuroscience* 8.12 (2005): 1704-1711.

[18] Dayan, Peter, Sham Kakade, and P. Read Montague. "Learning and selective attention." *nature neuroscience* 3 (2000): 1218-1223.

[19] D'Esposito, Mark, et al. "Maintenance versus manipulation of information held in working memory: an event-related fMRI study." *Brain and cognition* 41.1 (1999): 66-86.

[20] Dias, Ângelo. "Mechanistic characterization of reinforcement learning in healthy humans using computational models." (2014).

[21] Dickinson, Anthony. "Actions and habits: the development of behavioural autonomy." *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 308.1135 (1985): 67-78.

[22] Didelot, Xavier, et al. "Likelihood-free estimation of model evidence." *Bayesian analysis* 6.1 (2011): 49-76.

[23] Doya, Kenji. "Reinforcement learning: Computational theory and biological mechanisms." *Hfsp j* 1.1 (2007).

[24] Evenden, J. L., and T. W. Robbins. "Increased response switching, perseveration and perseverative switching following d-amphetamine in the rat."*Psychopharmacology* 80.1 (1983): 67-73.

[25] Foerde, Karin, and Daphna Shohamy. "The role of the basal ganglia in learning and memory: insight from Parkinson's disease." *Neurobiology of learning and memory* 96.4 (2011): 624-636.

[26] Frank, Michael J., Anouk Scheres, and Scott J. Sherman. "Understanding decision-making deficits in neurological conditions: insights from models of natural action selection." *Philosophical Transactions of the Royal Society B: Biological Sciences* 362.1485 (2007): 1641-1654.

[27] Frank, Michael J., Bryan Loughry, and Randall C. O'Reilly. "Interactions between frontal cortex and basal ganglia in working memory: a computational model." *Cognitive, Affective, & Behavioral Neuroscience* 1.2 (2001): 137-160.

[28] Frank, Michael J., et al. "Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning." *Proceedings of the National Academy of Sciences* 104.41 (2007): 16311-16316.

[29] Frank, Michael J., et al. "Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation." *Nature neuroscience* 12.8 (2009): 1062-1068.

[30] Frank, Michael J., Randall C. O'Reilly, and Tim Curran. "When memory fails, intuition reigns Midazolam enhances implicit inference in humans."Psychological Science 17.8 (2006): 700-707.

[31] Friston, Karl, et al. "Variational free energy and the Laplace approximation."*NeuroImage* 34.1 (2007): 220-234.

[32] Gluck, Mark A., Daphna Shohamy, and Catherine Myers. "How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning." *Learning & Memory* 9.6 (2002): 408-418.

[33] Henke, Katharina. "A model for memory systems based on processing modes rather than consciousness." *Nature Reviews Neuroscience* 11.7 (2010): 523-532.

[34] Jay, Thérèse M. "Dopamine: a potential substrate for synaptic plasticity and memory mechanisms." *Progress in neurobiology* 69.6 (2003): 375-390.

[35] Jog, Mandar S., et al. "Building neural representations of habits." *Science*286.5445 (1999): 1745-1749.

[36] Kakade, Sham, and Peter Dayan. "Acquisition and extinction in autoshaping."*Psychological review* 109.3 (2002): 533.

[37] Kéri, Szabolcs, et al. "Probabilistic classification learning in Tourette syndrome." *Neuropsychologia* 40.8 (2002): 1356-1362.
[38] Knowlton, Barbara J., Jennifer A. Mangels, and Larry R. Squire. "A neostriatal habit learning system in humans." *Science* 273.5280 (1996): 1399-1402.
[39] Lau, Brian, and Paul W. Glimcher. "DYNAMIC RESPONSE-BY-RESPONSE MODELS OF MATCHING BEHAVIOR IN RHESUS MONKEYS." *Journal of the experimental analysis of behavior* 84.3 (2005): 555-579.
[40] Leckman, James F., et al. "Pathogenesis of Tourette's syndrome." *Journal of Child Psychology and Psychiatry* 38.1 (1997): 119-142.
[41] Lilliefors, Hubert W. "On the Kolmogorov-Smirnov test for normality with mean and variance unknown." *Journal of the American Statistical Association* 62.318 (1967): 399-402.
[42] Ludvig, Elliot A., Marc G. Bellemare, and Keir G. Pearson. "A primer on reinforcement learning in the brain: Psychological, computational, and neural perspectives." *Computational neuroscience for advancing artificial intelligence: Models, methods and applications* (2011): 111-144.
[43] Maia, Tiago V., and Michael J. Frank. "From reinforcement learning models to psychiatric and neurological disorders." *Nature neuroscience* 14.2 (2011): 154-162.
[44] Maia, Tiago V. "Reinforcement learning, conditioning, and the brain: Successes and challenges." *Cognitive, Affective, & Behavioral Neuroscience* 9.4 (2009): 343-364.
[45] Marsh, Rachel, et al. "Habit learning in Tourette syndrome: a translational neuroscience approach to a developmental psychopathology." *Archives of general psychiatry* 61.12 (2004): 1259-1268.
[46] Mathews, Carol A., and Marco A. Grados. "Familiality of Tourette syndrome, obsessive-compulsive disorder, and attention-deficit/hyperactivity disorder: heritability analysis in a large sib-pair sample." *Journal of the American Academy of Child & Adolescent Psychiatry* 50.1 (2011): 46-54.
[47] Mehler-Wex, C., P. Riederer, and M. Gerlach. "Dopaminergic dysbalance in distinct basal ganglia neurocircuits: implications for the pathophysiology of Parkinson's disease, schizophrenia and attention deficit hyperactivity disorder."*Neurotoxicity research* 10.3-4 (2006): 167-179.
[48] Middleton, Frank A., and Peter L. Strick. "Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies." *Brain and cognition*42.2 (2000): 183-200.
[49] Milner, Brenda, Suzanne Corkin, and H-L. Teuber. "Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of HM."*Neuropsychologia* 6.3 (1968): 215-234.
[50] Montague, P. Read, Peter Dayan, and Terrence J. Sejnowski. "A framework for mesencephalic dopamine systems based on predictive Hebbian learning." *The Journal of neuroscience* 16.5 (1996): 1936-1947.
[51] Morris, Genela, et al. "Midbrain dopamine neurons encode decisions for future action." *Nature neuroscience* 9.8 (2006): 1057-1063.
[52] Nemoda, Zsofia, Anna Szekely, and Maria Sasvari-Szekely. "Psychopathological aspects of dopaminergic gene polymorphisms in adolescence and young adulthood." *Neuroscience & Biobehavioral Reviews*35.8 (2011): 1665-1686.
[53] Niv, Yael, et al. "Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain." *The Journal of Neuroscience* 32.2 (2012): 551-562.
[54] O'Reilly, Randall, and Michael Frank. "Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia."*Neural computation* 18.2 (2006): 283-328.
[55] Packard, Mark G., and Barbara J. Knowlton. "Learning and memory functions of the basal ganglia." *Annual review of neuroscience* 25.1 (2002): 563-593.
[56] Palminteri, Stefano, et al. "Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes." *Proceedings of the National Academy of Sciences* 106.45 (2009): 19179-19184.
[57] Palminteri, Stefano, et al. "Similar improvement of reward and punishment learning by serotonin reuptake inhibitors in obsessive-compulsive disorder."*Biological psychiatry* 72.3 (2012): 244-250.
[58] Pauls, David L., et al. "Obsessive-compulsive disorder: an integrative genetic and neurobiological perspective." *Nature Reviews Neuroscience* 15.6 (2014): 410-424.
[59] Penny, W. D. "Comparing dynamic causal models using AIC, BIC and free energy." *Neuroimage* 59.1 (2012): 319-330.
[60] Penny, Will D., et al. "Comparing dynamic causal models." *Neuroimage* 22.3 (2004): 1157-1172.
[61] Poldrack, Russell A., and Mark G. Packard. "Competition among multiple memory systems: converging evidence from animal and human brain studies."*Neuropsychologia* 41.3 (2003): 245-251.
[62] Poldrack, Russell A., and Paul Rodriguez. "How do memory systems interact? Evidence from human classification learning." *Neurobiology of learning and memory* 82.3 (2004): 324-332.
[63] Poldrack, Russell A., et al. "Interactive memory systems in the human brain."*Nature* 414.6863 (2001): 546-550.
[64] Price, Amanda L. "Distinguishing the contributions of implicit and explicit processes to performance of the weather prediction task." *Memory & cognition*37.2 (2009): 210-222.
[65] Reber, Paul J., Barbara J. Knowlton, and Larry R. Squire. "Dissociable properties of memory systems: differences in the flexibility of declarative and nondeclarative
[66] Ridley, Rosalind M., et al. "Stereotyped responding on a two-choice guessing task by marmosets and humans treated with amphetamine."*Psychopharmacology* 95.4 (1988): 560-564.
[67] Rigoux, Lionel, et al. "Bayesian model selection for group studies—Revisited."*Neuroimage* 84 (2014): 971-985.
[68] Roesch, Matthew R., Donna J. Calu, and Geoffrey Schoenbaum. "Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards." *Nature neuroscience* 10.12 (2007): 1615-1624.
[69] Rutledge, Robb B., et al. "Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task." *The Journal of Neuroscience* 29.48 (2009): 15104-15114.
[70] Schönberg, Tom, et al. "Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making."*The Journal of Neuroscience* 27.47 (2007): 12860-12867.
[71] Schultz, Wolfram, Peter Dayan, and P. Read Montague. "A neural substrate of prediction and reward." *Science* 275.5306 (1997): 1593-1599.
[72] Scoville, William Beecher, and Brenda Milner. "Loss of recent memory after bilateral hippocampal lesions." *Journal of neurology, neurosurgery, and psychiatry* 20.1 (1957): 11.
[73] Shohamy, D., et al. "Basal ganglia and dopamine contributions to probabilistic category learning." *Neuroscience & Biobehavioral Reviews* 32.2 (2008): 219-236.

[74] Shohamy, D., et al. "Role of the basal ganglia in category learning: how do patients with Parkinson's disease learn?." *Behavioral neuroscience* 118.4 (2004): 676.

[75] Squire, Larry R. "Memory systems of the brain: a brief history and current perspective." *Neurobiology of learning and memory* 82.3 (2004): 171-177.

[76] Stephan, Klaas Enno, et al. "Bayesian model selection for group studies."*Neuroimage* 46.4 (2009): 1004-1017.

[77] Sutton, Richard S., and Andrew G. Barto. *Introduction to reinforcement learning*. MIT Press, 1998.

[78] Tricomi, Elizabeth, Bernard W. Balleine, and John P. O'Doherty. "A specific role for posterior dorsolateral striatum in human habit learning." *European Journal of Neuroscience* 29.11 (2009): 2225-2232.

[79] Walitza, Susanne, et al. "Transmission disequilibrium studies in early onset of obsessive–compulsive disorder for polymorphisms in genes of the dopaminergic system." *Journal of neural transmission* 115.7 (2008): 1071-1078.

[80] Wiecki, Thomas V. "Sequential sampling models in computational psychiatry: Bayesian parameter estimation, model selection and classification." *arXiv preprint arXiv:1303.5616* (2013).

[81] Worbe, Yulia, et al. "Reinforcement learning and Gilles de la Tourette syndrome: dissociation of clinical phenotypes and pharmacological treatments." *Archives of general psychiatry* 68.12 (2011): 1257-1266.

[82] Yin, Henry H., and Barbara J. Knowlton. "The role of the basal ganglia in habit formation." *Nature Reviews Neuroscience* 7.6 (2006): 464-476.