

Valence-dependent dopaminergic modulation during reversal learning in Parkinson's disease: A neurocomputational approach

Mauro Ursino^{a,*}, Silvana Pelle^a, Fahima Nekka^{b,d,e}, Philippe Robaey^c, Miriam Schirru^{a,b}

^a Department of Electrical, Electronic and Information Engineering Guglielmo Marconi, University of Bologna, Campus of Cesena, I 47521 Cesena, Italy

^b Faculté de Pharmacie, Université de Montréal, Montreal, Quebec H3T 1J4, Canada

^c Children's Hospital of Eastern Ontario, University of Ottawa, Ottawa, ON, Canada

^d Centre de recherches mathématiques, Université de Montréal, Montreal, Quebec H3T 1J4, Canada

^e Centre for Applied Mathematics in Bioscience and Medicine (CAMBAM), McGill University, Montreal, Quebec H3G 1Y6, Canada

ARTICLE INFO

Keywords:

Dopamine
Reward
Punishment
Hebb rule
Reinforcement learning
Cognitive Flexibility
Go NoGo pathways

ABSTRACT

Reinforcement learning, crucial for behavior in dynamic environments, is driven by rewards and punishments, modulated by dopamine (DA) changes. This study explores the dopaminergic system's influence on learning, particularly in Parkinson's disease (PD), where medication leads to impaired adaptability. Highlighting the role of tonic DA in signaling the valence of actions, this research investigates how DA affects response vigor and decision-making in PD. DA not only influences reward and punishment learning but also indicates the cognitive effort level and risk propensity in actions, which are essential for understanding and managing PD symptoms.

In this work, we adapt our existing neurocomputational model of basal ganglia (BG) to simulate two reversal learning tasks proposed by Cools et al. We first optimized a Hebb rule for both probabilistic and deterministic reversal learning, conducted a sensitivity analysis (SA) on parameters related to DA effect, and compared performances between three groups: PD-ON, PD-OFF, and control subjects.

In our deterministic task simulation, we explored switch error rates after unexpected task switches and found a U-shaped relationship between tonic DA levels and switch error frequency. Through SA, we classify these three groups. Then, assuming that the valence of the stimulus affects the tonic levels of DA, we were able to reproduce the results by Cools et al.

As for the probabilistic task simulation, our results are in line with clinical data, showing similar trends with PD-ON, characterized by higher tonic DA levels that are correlated with increased difficulty in both acquisition and reversal tasks.

Our study proposes a new hypothesis: valence, signaled by tonic DA levels, influences learning in PD, confirming the uncorrelation between phasic and tonic DA changes. This hypothesis challenges existing paradigms and opens new avenues for understanding cognitive processes in PD, particularly in reversal learning tasks.

1. Introduction

When a subject interacts with the environment without explicit instructions, learning is driven by unexpected rewards and punishments (reinforcement learning). It is well known that phasic dopamine (DA) changes (peaks and dips), which encode the difference between the expected and the actual outcome, drive this process (Schultz, 1998). The dopaminergic system not only works to find the optimal behavior in a given context but also enables flexibility (i.e., reversal learning) in the

presence of a non-stationary environment when positive and negative contingencies can change.

Significant human neurological or psychiatric disorders, such as Parkinson's disease (PD), schizophrenia, Attention Deficit Hyperactivity Disorder (ADHD), depression, addiction, and post-traumatic stress disorders, implicate dysfunction of the DA system and, consequently, abnormal behavior. Besides their typical symptoms, these disorders are often characterized by the difficulty of modifying one's choices despite adverse consequences (Grace, 2016; Klein et al., 2019).

* Corresponding author at: Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi", Area di Campus Cesena, Via Dell'Università 50, I 47521 Cesena FC, Italy.

E-mail addresses: mauro.ursino@unibo.it (M. Ursino), silvana.pelle3@unibo.it (S. Pelle), fahima.nekka@umontreal.ca (F. Nekka), philippe.robaey@umontreal.ca (P. Robaey), miriam.schirru@umontreal.ca (M. Schirru).

<https://doi.org/10.1016/j.nlm.2024.107985>

Received 22 March 2024; Received in revised form 19 August 2024; Accepted 6 September 2024

Available online 11 September 2024

1074-7427/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

A pivotal role in reinforcement learning is played by the basal ganglia (BG), a subcortical structure implicated in action selection. In fact, DA changes affect synapse training in the striatum (Hebbian potentiation and depotentiation) through D1 and D2 receptors and establish a different weight for the Go and NoGo pathways during learning (Gerfen and Surmeier, 2011). This is essential to favor or avoid specific actions.

Despite the enormous number of experimental, clinical, and theoretical studies on the subject in recent years, the relationship between the dopaminergic system, the BG, and reversal learning is still under active investigation, with many points requiring additional analysis. Understanding these relationships is vital for treating patients with the aforementioned neurological disorders and optimizing the current therapies.

In the following, we will focus on reversal learning in PD patients since many experimental and theoretical studies are concerned with this pathology; moreover, these patients have been studied both on levodopa medication and without medication, thus emphasizing the role of DA changes. Interestingly, PD patients frequently develop severely disabling side effects after levodopa medication in the form of impulse control disorder, psychosis, and addiction (Dagher and Robbins, 2009; Driver-Dunckley et al., 2003; Lawrence et al., 2003; Voon et al., 2009). In particular, ON-medicated patients often exhibit a reduced ability in reversal learning compared with OFF-medicated patients and control subjects, both during deterministic and probabilistic cognitive tasks (Cools et al., 2006, 2001; Frank et al., 2004; Swainson et al., 2000).

To understand the mechanisms behind these cognitive side effects, Cools et al. (Cools et al., 2022) developed a hypothesis named “dopamine overdose”; this is based on the observation that the substantia nigra pars compacta projects primarily to the dorsal rather than to the ventral striatum (Kish et al., 1988; Rinne, 1993). Consequently, levodopa treatment in PD patients (at least at the beginning of the disease) can restore normal DA levels in the dorsal part, mainly implicated in motor control. However, this can also lead to a DA excess in the ventral part, which is especially crucial for cognitive decision-making (Rinne, 1993). While the dorsal region, experiencing DA depletion, benefits from levodopa supplementation for restoration, the ventral part is less affected by neuron loss and thus becomes sensitive to DA overdose. This can be explained by the selective denervation process, which is more prominent in the dorsal part than in the ventral part (Fearnley and Lees, 1991).

A consequential hypothesis is that, due to low basic DA levels, OFF-medicated PD patients should have difficulties in choosing rewarded actions. In contrast, ON-medicated PD patients should be less sensitive to punishments due to higher DA levels in the ventral striatum. Since reversal learning is primarily based on uncorrected choices (at least during the first phase of reversal after an environment shift), it is markedly impacted in subjects with high DA levels.

Several results in the literature support this scenario, suggesting that ON-medicated PD patients exhibit better reward-based learning (Rutledge et al., 2009) but impaired punishment-based (reversal) learning compared to OFF-medicated patients (Bódi et al., 2009; Cools et al., 2006; Frank, 2006; Frank et al., 2004; Graef et al., 2010; McCoy et al., 2019; Moustafa et al., 2008). Other studies, however, failed to find similar differences (Coulthard et al., 2012; Grogan et al., 2017). Generally, a large individual variability can manifest in the effects, and the mechanisms underlying them often remain unclear.

All these aspects have been summarized in several recent neurocomputational models, which offer essential insights into reinforcement learning in the BG and provide a unifying account of the main mechanisms involved (Cohen and Frank, 2009; Cutsuridis and Perantonis, 2006; Frank, 2005; Humphries et al., 2018; Kato and Morita, 2016; Moustafa et al., 2014; Moustafa and Gluck, 2011; Schroll and Hamker, 2013; Véronneau-Veilleux et al., 2021). According to a classic schema (Gerfen and Surmeier, 2011), individual actions are represented in the BG through segregated channels, each characterized by its Go (direct)

and NoGo (indirect) pathways. DA is excitatory on the Go pathway, which facilitates the response, and inhibitory on the NoGo pathway, which inhibits the response. This theoretical underpinning has been used in most models and can explain the main aspects of medication in PD patients (Baston et al., 2016; Frank et al., 2004; Frank, 2005; Ursino et al., 2020a) and in different pathologies such as ADHD (Véronneau-Veilleux et al., 2022) and Huntington’s disease (Schroll et al., 2015).

Despite this general schema being well accepted and confirmed by experimental and theoretical studies, several aspects still deserve clarification, especially regarding reversal learning. During reversal learning, training after each switch (when the association between a stimulus and a rewarded action changes) is primarily driven by unexpected errors, which, according to the previous schema, should be associated with a DA dip and reinforcement of the NoGo pathway. In contrast, the subsequent consolidation (when the subject has learned the new association and performs correctly) is primarily driven by rewards, which should further potentiate the Go pathway. Reversal learning tasks, in turn, can be deterministic, involving fast learning, or probabilistic, where rewards and punishments are more balanced according to a given statistic.

Within this scenario, we have identified several issues that can benefit from further neurocomputational analysis, as detailed in the following:

(i) What is the best form of the Hebb rule for synapse potentiation/depotentialization in the striatum that is able to simulate different tasks (deterministic and probabilistic) within a single unifying theoretical framework?

(ii) What is the specific role of the Go and NoGo pathway? A recent review (Calabresi et al., 2014) suggests that these two pathways should not be considered independent but functionally correlated in a push-pull manner.

(iii) What is the specific role of tonic DA vs. phasic DA? Do pathological aspects and the differences between medicated vs. non-medicated patients mainly result from alterations in tonic DA, phasic DA, or both?

(iv) Several recent studies suggest that tonic DA not only affects the capacity to learn from unexpected rewards or unexpected punishments but also signals the valence of a given action and/or implement a conditioned behavior (Niv et al., 2007; Rigoli et al., 2016a; Saunders et al., 2018). In this context, valence refers to the positive or negative nature of the outcome. How do DA levels (or other neuromodulators, such as norepinephrine or serotonin (Dayan and Huys, 2008)) reflect this aspect of valence on task performance in the BG?

(v) It is well-accepted that phasic DA changes are an indicator of an unexpected outcome (unexpected reward or punishment). How does the BG or the prefrontal cortex memorize this expectancy based on a previous history of rewards/punishments? Many models use the present reward or punishment without explicitly caring about expectations.

To address these points, we modified our previous model (Schirru et al., 2022) to simulate two classic reversal learning tasks, the deterministic one (Cools et al., 2009, 2006) and the probabilistic one (Cools et al., 2001). Simulation of the deterministic task was essential to finding an adequate Hebb rule, allowing us to learn in just one step and disentangle a possible role for the Go and NoGo pathways. To understand the impact of DA-related parameters, a sensitivity analysis (SA) was performed, and differences between medicated and non-medicated patients were tested and compared with the performance of control subjects in both trials, providing additional insights for points iii), iv), and v). Finally, the original aspects of this work, possible testable predictions, and lines for future improvement are discussed.

2. Methods

2.1. Main model assumptions

The model makes use of several fundamental assumptions. Some of

them have already been exploited in previous works, where a detailed justification can be found (Baston and Ursino, 2015; Ursino and Baston, 2018): (i) each action is characterized by a segregated channel, with its own Go and NoGo pathways, characterized by different DA responses; (ii) the different actions compete in the motor cortex, according to a “Winner takes all” arrangement of lateral synapses; (iii) a re-entrant thalamic connection, disinhibited by the Striatum, is necessary to determine the winner and to start the action; (iv) the hyper-direct pathway has the function of solving conflicts within the motor cortex.

The present work introduces a set of crucial new assumptions, which not only bring novelty but also significantly reshape the model’s functionality and outcomes. These assumptions stand as the key differentiators from other neurocomputational models:

- i) Before each choice, the value of the tonic DA input is affected by the valence of the expected result: an emotionally positive expectation may cause a tonic DA increase, whereas an emotionally negative expectation may cause a tonic DA decrease. Various results in the literature support this assumption, showing that tonic DA level signals the valence of an action (Niv et al., 2007; Rigoli et al., 2016b; Zénon et al., 2016). Additionally, Morita and Kato (2022) highlighted the importance of a recent study by Mikhael et al. (2022), which showed that DA signals, representing reward prediction errors, slightly ramp towards reward timings. These errors are used for accurate value learning in conditions with uncertainty about upcoming states, resolved by sensory feedback. This DA signal characteristic provides further validation for our assumption regarding anticipatory DA changes. Furthermore, a recent study (Delignat-Lavaud et al., 2023) demonstrated that even when activity-dependent phasic DA release is reduced by 95 % in mice, behaviors that depend on DA remain unchanged or are even better, underscoring the critical role of tonic DA as emphasized by our modeling results.
- ii) The probability that a given choice is rewarded is coded by the activity of the winner Go neuron in the striatum (normalized between 0 and 1). We have not yet found a direct experimental confirmation of this idea, which, to our knowledge, has been partially exploited only by (Humphries et al., 2012); these authors assumed that the output of the BG could represent a probability distribution for action selection.
- iii) The relationship between phasic DA and reward expectation is highly non-linear. Phasic changes in DA (positive or negative) become increasingly significant as the result deviates more from the expectation. This is a crucial assumption supported by numerous studies, which demonstrate that most DA neurons are activated by a higher reward than predicted (positive prediction error) and depressed by a lower reward than predicted (negative prediction error) (Diederen et al., 2017; Schultz, 2016, 1998; Schultz et al., 1997). Our choice differs from that of most previous models. Some authors used contrastive Hebbian learning (CHL), computing the difference of the Hebbian product (pre and postsynaptic activation product) across two states, namely, the network’s actual output phase and a subsequent phase in which the target output is experienced (Frank, 2006; Frank and Claus, 2006). Other models make use of the classic Sutton-Barton algorithm to compute a temporal difference error signal and use this signal as a multiplicative factor in the Hebb rule (Moustafa et al., 2014). More complex models involve an actor-critic system, where the critic is a Pavlovian learning system that controls the firing of simulated midbrain DA neurons and trains both itself and the actor, i.e., the BG (O’Reilly and Frank, 2006).
- iv) Learning is governed by a complete Hebb rule in the striatum, including LTP and LTD. The rule applies to all possible combinations of pre-synaptic and post-synaptic activities. Possible synapse plasticity in other regions of the BG or of the cortex, although documented in the literature, is not essential for the

present results. The possibility of this version of the Hebb rule is supported by the presence of inhibitor interneurons in the striatum, making a disynaptic connection between input and output neurons (Di Filippo et al., 2009; Fino and Venance, 2011).

2.2. Model description

Qualitative model description – The model implemented is a representation of the human BG based on our previous work (Baston and Ursino, 2015; Schirru et al., 2022). A schematic diagram of the model is presented in Fig. 1.

The whole network comprises several neural units. Each unit (simulating a group of neurons with similar functions) is characterized by first-order low-pass dynamics to reproduce the integrative property of the membrane and a sigmoidal relation for output activity to represent the presence of lower and upper saturation values for neuronal activity.

The model comprises a sensory representation (S) and a motor representation (C) in the cortex. The sensory representation’s neurons correspond to the stimulus presented to the network, while the motor neurons in the cortex encode possible actions. In addition to the previous neural units downstream the cortex, the model includes the striatum, further subdivided into Go and NoGo pathways, the globus pallidus pars externa (Gpe), the globus pallidus pars interna (Gpi), the thalamus (T), all with a neuron count in a 1-to-1 relationship with cortex, the cholinergic interneuron (ChI) and the subthalamic nucleus (STN), modeled as single neurons representing the entire population activity.

The model integrates three primary pathways: direct (via Go neurons), indirect (via NoGo neurons), and hyperdirect (via STN). The combined action of these pathways inhibits the thalamus from Gpi, either permitting or blocking a coded response by motor neurons. In the absence of sufficient stimulation, the network maintains a basal steady-state, with inhibited cortex, striatum, and thalamus, and basal activity in Gpi and Gpe, in accordance with physiological data (van Albada and Robinson, 2009).

Once a sufficient stimulus is provided to the network, the motor cortex can select a response based on competition among neurons, implemented by a winner-takes-all (WTA) mechanism and featuring lateral inhibition and an excitatory self-loop. After action selection, both direct and indirect paths operate in parallel for each neuron (Mink, 1996). Specifically, each neuron of the cortex is connected to its corresponding neuron in the Go and NoGo pathways via excitatory synapses. The Go pathway facilitates the response by direct inhibition of the Gpi, resulting in thalamus disinhibition. In contrast, the NoGo pathway blocks the response by inhibiting Gpe, which sends inhibitory synapses to Gpi, thus further inhibiting the thalamus. The hyperdirect pathway comes into play whenever intense conflict arises among different neurons in the cortex, i.e., two or more neurons fire together despite the presence of a winner-takes-all mechanism, necessitating additional time for the cortex to select the winning neuron. The STN, receiving the conflict level as input, sends excitation to Gpi, resulting in thalamus inhibition in this scenario.

Dopamine role – As mentioned in the introduction, DA plays a pivotal role in learning mechanisms, with its effect either excitatory or inhibitory, depending on the type of receptors it binds to (excitatory when binding to D1 receptors and inhibitory when D2 receptors are involved). In the model, this distinction translates into different DA effects on the Go pathway (predominantly excitatory) and the NoGo pathway (primarily inhibitory).

At the basal steady-state level, DA maintains a tonic level. In case of reward or punishment, a phasic change in DA occurs. In the case of rewards, DA induces excitation on the active Go neuron and inhibition on inactive Go neurons, with a contrast enhancement mechanism to favor only one neuron winning. Simultaneously, it exerts an inhibitory effect on all NoGo neurons. In the case of punishments, all NoGo neurons are excited and Go neurons are inhibited. Compared with our previous

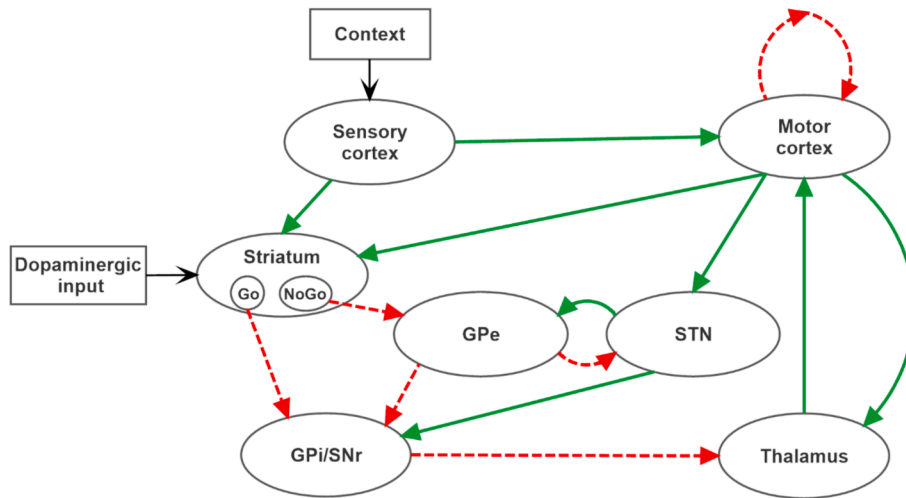


Fig. 1. Block diagram describing the primary regions involved in the BG model and their relationships. Continuous green lines represent excitatory synapses, and dashed red lines represent inhibitory synapses. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

works (Baston and Ursino, 2015; Schirru et al., 2022; Ursino and Baston, 2018), we here assumed that the excitatory effect of a DA dip on the NoGo neurons exhibits a maximal value (see Eq.10 in Supplementary Materials 1). This is important to prevent a scenario where all NoGo neurons are excessively excited following an unexpected punishment, which could result in all actions being punished. To address this, saturation has been implemented. This ensures that reinforcement is only directed to the NoGo pathway associated with the action unexpectedly punished, reducing the repetition of this particular action under similar future conditions.

In addition, cholinergic interneurons amplify the phasic DA effect through a push-pull mechanism. In detail, cholinergic interneurons are capable of sensing DA changes: DA drops increase Chl activity, resulting in disinhibition of the Go pathways and inhibition of NoGo pathways. The opposite holds in the case of DA peaks.

In order to reach a better understanding of reward and punishment mechanisms in ventral BG and emphasize the role of phasic and tonic DA changes, we examine two aspects of the model in detail in the following section. All other equations and parameter numerical values can be found in Supplementary Materials 1.

(i) What is the most appropriate form of learning rule for striatal synapses that can be used to simulate both deterministic and probabilistic learning tasks within a unified theoretical framework?

(ii) How do previous experiences influence DA changes, translated into larger changes for unexpected cases (punishments or rewards) and smaller changes for expected ones?

Hebbian rule – In a recent study, we compared several forms of Hebbian rules to simulate reversal learning during two-choice or four-choice probabilistic tasks (Schirru et al., 2022). The best rule was one based on the post-synaptic activity of striatal neurons, i.e., striatal synapses were modified only if the corresponding post-synaptic Go or NoGo neurons were active.

However, we found the same rule inadequate to simulate a deterministic task (as in (Cools et al., 2009, 2006) in which a subject must modify his response immediately after a switch) or a single-choice probabilistic task (as in (Cools et al., 2001; Swainson et al., 2000)). In the following, we will examine the behavior of three alternative rules simulating Cools et al. deterministic experiments. We remark that, in each case, we introduced upper and lower saturation values (named w_{max} and w_{min} in Table 1 of Supplementary Materials 1) for each trained synapse during learning.

The fundamental idea is that the Hebb rule is based on comparing neuron activity with a threshold to determine activation or inactivation

(see Eqs. 1–3 below). Basically, we postulate that after a reward, only one winner Go neuron is excited, whereas the other Go neuron and all NoGo neurons are inhibited. Conversely, after an unexpected punishment, all Go neurons are inhibited (due to a phasic DA fall). In contrast, the NoGo neuron in the winner channel is active, and the other NoGo neuron activity is below the threshold.

The three Hebb rules examined hereafter are:

The post-synaptic rule

$$\Delta w_{ij}^{AB} = \sigma(y_j^B - \vartheta^{PRE})(y_i^A - \vartheta^{POST})^+ \quad (1)$$

where Δw_{ij}^{AB} is the variation of the synapse between the pre-synaptic neuron j in layer B (B=S or C) and the post-synaptic neuron i , in layer A (A=G or N), y represents neuron activity (normalized between 0 and 1), ϑ^{PRE} and ϑ^{POST} are thresholds for the pre-synaptic and post-synaptic activity, and the expression $()^+$ represents the function *positive part* (i.e., $(u)^+ = u$ if $u > 0$, 0 otherwise). According to Eq. (1), the rule is applied only if the post-synaptic neuron activity is above the threshold.

The ex-or rule

$$\Delta w_{ij}^{AB} = \begin{cases} \sigma(y_j^B - \vartheta^{PRE})(y_i^A - \vartheta^{POST}) & \text{if } y_j^B > \vartheta^{PRE} \text{ or } y_i^A > \vartheta^{POST} \\ 0 & \text{if } y_j^B \leq \vartheta^{PRE} \text{ and } y_i^A \leq \vartheta^{POST} \end{cases} \quad (2)$$

This rule excludes only the case when both post-synaptic and pre-synaptic activities are below threshold.

The complete rule

$$\Delta w_{ij}^{AB} = \sigma(y_j^B - \vartheta^{PRE})(y_i^A - \vartheta^{POST}) \quad (3)$$

This rule considers all possible cases, including a synapse reinforcement when both pre-synaptic and post-synaptic neurons are below the threshold. The possibility of a reinforcement when both neurons are below the threshold may appear unphysiological. Still, it can be justified by a disinhibition (see Supplementary Materials 2), i.e., assuming a decrease in an excitatory synapse from the input neuron to an intermediate inhibitory interneuron (see Discussion). The presence of inhibitory interneurons in the Striatum is well documented (Di Filippo et al., 2009; Fino and Venance, 2011).

To evaluate the three Hebb rules, let us assume that the network just learned to associate an Action1 to a first stimulus (S=[1 0] in Fig. 2 and Figs S2 and S3 in Supplementary Materials 2) and an Action 2 to a second stimulus (S=[0 1] in the same figures). At a given moment, a switch occurs; for instance, the second stimulus is now rewarded after

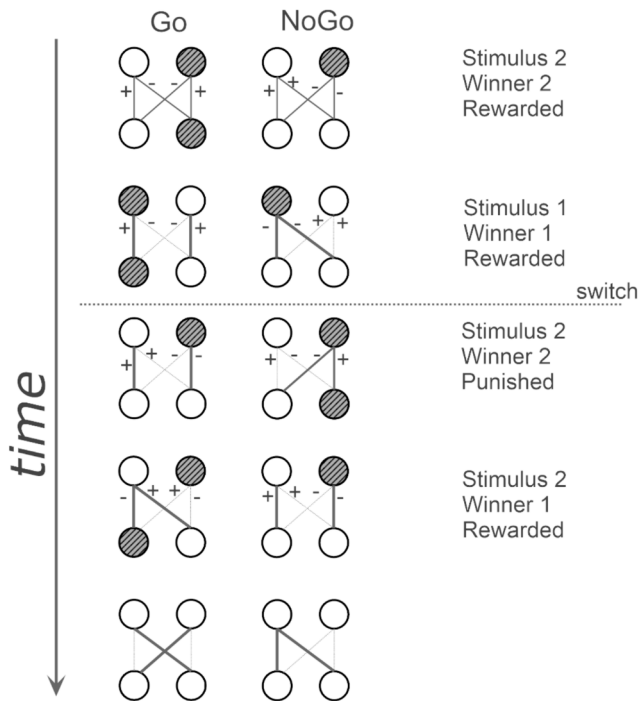


Fig. 2. Qualitative simulation of a deterministic task (as in Cools et al., 2006) assuming two action channels and using the complete Hebb rule (Eq. (3)). Each panel represents the synapses connecting the external stimulus S (in the present model $S=[1\ 0]$ for the first stimulus and $S=[0\ 1]$ for the second stimulus) with the two neurons in the Go pathways (left panels) and the two neurons in the NoGo pathways (right panels). Both the Go and NoGo exhibit two neurons, representing the first choice and the second choice, respectively. A white circle represents a silent neuron (activity below the threshold), while a dashed circle represents an excited neuron (above the threshold). Three possible values are assumed for the synapses (low: thin line; intermediate: medium line; high: thick line); these values are the result of previous learning. After each choice, synapses are potentiated (sign +) or depotentiated (sign -) according to Eq. (3). Moreover, we assume that the learning rate is so strong as to modify the synapse value directly from the low to the high value (in case of potentiation) or from the high to the low value (in case of depotentiation) in just a single trial. This holds for the deterministic task only. In the first two rows (before a switch), we assume that the first stimulus is rewarded by Action 1, and the second stimulus is rewarded by Action 2. In the first row, we assume that starting from a naive condition (all synapses have an intermediate value), the participant receives Stimulus 2, responds casually with Action 2, and is rewarded (this is the unique initial random choice; all the subsequent choices are deterministic). After a reward, only the Go neuron in the winner pathway is excited; all other neurons are inhibited. In the second row, the subject receives Stimulus 1, responds with Action 1 (according to the present new value of Go synapses), and is rewarded. At the switch moment, the Go determines the correct choices, while the NoGo is ambiguous. From the third row downward, we simulate a reversal (now Stimulus 1 is rewarded by Action 2, and Stimulus 2 is rewarded by Action 1). In the third row, the subject receives the second Stimulus, responds with Action 2 (according to the present Go synapses), and is punished. After a punishment, only the NoGo neuron in the winner pathway is excited; all other neurons are inhibited. In the fourth row, the subject receives the Stimulus 2 again (as in Cools et al., 2006). But now the NoGo dominates (while the Go is ambiguous) and inhibits the wrong choice. Hence, the subject correctly responds with Action 1 and is rewarded. After this reward, the Go dominates again (fifth row, note the reversal of synapses) and signals the correct choice, while the NoGo is ambiguous.

Action 1 and punished after Action 2. Since the network erroneously responds with Action 2 (due to previous learning), an unexpected punishment has now occurred.

Immediately after the switch, the second stimulus is presented again (that is, the same stimulus on which an error was done immediately before) to test whether the network was instantly able to learn the new

association (i.e., to respond with Action 1 instead of Action 2).

Results are summarized in the panels of Fig. 2 (for the case of the complete Hebb rule, Eq. (3)) and Figs. S1 and S2 of Supplementary Materials 2 (for what concerns the other two rules). These figures show that only the complete Hebb rule is able to ensure a correct shift after a single unexpected punishment. In contrast, the use of the other two rules results in an uncertain decision immediately after an unexpected switch. In other words, the subject can become unable to make a choice, producing a large number of switch errors. The qualitative results illustrated in these figures have been further supported by simulations (results not shown) performed on the overall model using the three alternative rules. Simulations confirm that only a complete Hebb rule correctly mimics the deterministic switch experiment, producing percentage errors of the order of those observed.

An interesting aspect emerging from Fig. 2 is that, after rewards, the Go portion of the BG dominates the behavior, becoming able to associate the correct stimulus with the correct response. At the same time, the NoGo withdraws its inhibition from the latter rewarded choice. However, after an unexpected punishment, the NoGo portion takes control. It dominates the response, becoming able to inhibit incorrect choices, whereas the Go portion withdraws excitation from the last punished action. Hence, both portions of the BG are essential to determine the correct behavior. The Go pathway comes into play after rewards, and the NoGo after punishments.

Phasic dopamine changes – In recent work (Schirru et al., 2022), we proposed an original mechanism to compute the effect of phasic DA changes on striatal neurons (in the following, this effect will be named ΔD) able to produce: (i) a high phasic peak effect after an unexpected reward but a negligible peak effect following an expected reward; (ii) a robust phasic dip effect following an unexpected punishment, but a negligible dip effect after a totally expected punishment.

Eqs. (4)–(6) below summarize the mechanism. It is worth noting that the quantity D does not represent DA concentration but rather the effect that DA can have on glutamatergic signaling in medium spiny neurons, which are projection neurons of the striatum. D values depend on D1 and D2 receptor sensitivity, reflecting how DA can influence the excitation of striatal (Go and NoGo) neurons. Therefore, this quantity can assume positive and negative values.

Equation (4) is based on the idea that the “expected reward” is signaled by the activity (normalized between 0 and 1) of the Go neuron in the winner channel when a reward or a punishment is given (i.e., just before the computation of phasic DA changes). An activity close to 1 signified a well-expected reward, and an activity close to 0.5 or even below was an unexpected reward.

Accordingly, we compute a reward expectancy (i.e., the estimated probability of a reward, noted $r_{expected}$) as follows:

$$r_{expected} = gO_w(t_{response}) \quad (4)$$

where w represents the winner channel (in this work, $w = 1$ or 2), and $t_{response}$ is the instant at which a reward/punishment is given.

In the case of a reward, the phasic D peak ΔD_{reward} is a non-linear function of the difference between 1 and $r_{expected}$. The higher this difference, the higher the D peak and vice versa. We can write:

$$\Delta D_{reward} = [2^*(1 - r_{expected})]^m \cdot D_p \quad (5)$$

where m represents a parameter greater than 1, chosen empirically, and D_p is a multiplicative factor that sets the strength of the response. As it is clear from Equation (5), when the expected reward probability ($r_{expected}$) is 0.5, the phasic D peak is equal to D_p . If $r_{expected}$ is less than 0.5 (an unlikely situation for a winner), we have a much stronger phasic peak. If $r_{expected}$ is close to 1 (in case of a strongly expected reward), the phasic D peak decreases dramatically to zero. The higher the m , the stronger the difference between an unexpected and an expected reward.

In the case of punishment, the higher the expected reward, the

stronger the phasic D drop $\Delta D_{\text{punishment}}$ (i.e., an unexpected punishment causes a substantial dip). We can write:

$$\Delta D_{\text{punishment}} = -0.5 \cdot [2 \cdot r_{\text{expected}}]^m \cdot D_p \quad (6)$$

If the expected reward is 0.5, the phasic D drop change is equal to $-D_p/2$. If r_{expected} increases, the phasic dopaminergic drop is dramatically amplified, thus automatically implementing the significant sensitivity to unexpected punishments.

Examples of phasic D changes, with different values of expected reward and different values of parameter m , are shown in Fig. 3.

It is worth noting that, in previous work (Schirru et al., 2022), we used the same value for parameter D_p as the tonic DA level (i.e., $D_p = D_t$). The implicit assumption was that the higher the tonic DA level, the higher the phasic response, and vice versa. Conversely, in the present work, we use different values for D_t and D_p since various authors suggest that phasic and tonic DA changes are probably uncorrelated or inversely correlated (see (Grace, 2016, 2001, 1991)), i.e., no precise data is showing that large level of tonic DA should correspond to proportionally large levels of DA dips or peaks. The two parameters (D_t and D_p) will be the subject of a separate SA.

2.3. Task descriptions

2.3.1. Deterministic task

In the original experiment presented by Cools et al. (2006), participants (PD-OFF, PD-ON, Control) predicted outcomes in a simulated card game (Fig. 4). Participants were instructed to imagine themselves as a casino boss observing a player during a card game (see Cools et al. (2006) for more details).

Participants were presented with two images per trial; one of them is highlighted. By pressing a corresponding button, they had to predict the outcome of the player game: a win (green button) or loss (red button) associated with the highlighted image. This setup allows participants to learn these associations over time. During the task, reversal phases were introduced, where the previously learned associations were switched. Once the reversal is introduced, the image previously associated with a win represents a loss and vice versa. This tests the subject's adaptability and learning under changed conditions.

Crucially, the participant predicts "a posteriori" whether the player has won or lost, meaning they could not influence the results but only predict what happened.

The task consists of separate blocks, each comprising 120 trials, during which a reversal occurs several times: more precisely, a reversal takes place once an appropriate amount of knowledge is achieved. The criterion of knowledge is a predefined number of consecutive correct responses (ranging from 5 to 9, selected randomly) to prevent the predictability of the reversal.

Two different block types have been designed, which differ as to valence conditions. Subjects were not made aware of this difference. In the first block type, the participant must always learn the switch

following an unexpected win ("unexpected win block"); in the second block type, the participant always learns after an unexpected loss ("unexpected loss block"). Here, valence refers to the positive or negative nature of these outcomes (unexpected wins or unexpected losses).

In most previous studies, unexpected wins for the player are confounded with positive feedbacks for the participant (hence DA peaks), and unexpected losses for the player are confounded with negative feedbacks for the participant (DA dips). Implicitly, this assumes that wins are associated with the Go pathway and losses to the NoGo pathway of a *single action choice* (represented in Fig. 5a in terms of phasic DA changes).

In our opinion, however, making a parallel between the player's wins or losses and the DA peaks or dips in the participant is confusing, as the objective of the participant is not to win from the game but to make correct predictions. Hence, we speculate that DA dips probably occur after any wrong predictions made by the task participant, which can occur either after an unexpected loss or an unexpected win by the player. In other terms, predicting a win or predicting a loss are two separate actions associated with two channels in the BG, each with its own Go and NoGo pathways. Our following assumptions on DA changes are summarized in Fig. 5 b and c.

This distinction between reward/punishment prediction and the control of action is clearly outlined by Robinson et al. (Robinson et al., 2010), who used a similar task. The authors explicitly speculate about two possible strategies to distinguish between appetitive or aversive predictions: i) a strategy in which participants work primarily toward the win-associated action (Go) and treat the other as an alternative (NoGo); ii) a strategy in which the participants exhibit a different bias in response to two possible actions, one for win predictions and the other for loss predictions.

Most previous papers implicitly assume the strategy i), treating wins as rewards and losses as punishments (Fig. 5a). Conversely, in the present paper, we assume the strategy ii), making use of two distinct action channels and treating unexpected outcomes always as a DA dip (Fig. 5b and c).

To emphasize this point, in the first step, we simulated a hypothetical task with two choices but no different valence (i.e., no win or loss). The participant must only predict whether the choice is correct or not. With this approach, the focus is solely on the stimulus-response association. Reversal can be achieved by presenting unexpected feedback on either action. This task with positive/negative feedback but no win/loss is called "neutral" and is represented by the blue line in Fig. 5b and c.

In a second step, we added a valence to each choice (win/loss) from the observer's point of view. For the observer, an unexpected win/loss is always an error, represented by a phasic decrease in DA. First, we assumed that the valence (win or loss) affects the phasic DA dip, greater for an unexpected win and smaller for an unexpected loss, compared with the neutral condition (no win/loss). As we will see in the results section, this assumption does not produce satisfactory results. Hence, we introduced an alternative assumption (Fig. 5c): the "win prediction" is signaled by an anticipatory increase in DA level, or more generally, a change in the baseline level of DA in the striatum, as suggested in the literature (Dayan and Huys, 2008; Niv et al., 2007; Rigoli et al., 2016a). Conversely, the "loss prediction" is signaled by an anticipatory decrease in DA levels. The effect of valence is represented by the green and red lines in Fig. 5c.

In our model, we assume that the network initiates from a completely naïve state, wherein all synapses begin with equivalent values, and neither of the two actions is initially preferred. The presence or the absence of the stimulus holds identical significance. To further clarify, the synapses from the sensory cortex neurons to the Go and NoGo neurons, as well as those from the motor cortex neurons to the Go and NoGo neurons, are set to the value of 0.5. In addition, to ensure an appropriate level of exploration by the network, Gaussian white noise with zero mean value and standard deviation (SD) of 0.08 is applied as input to the neurons of the motor cortex. We have chosen a low value of

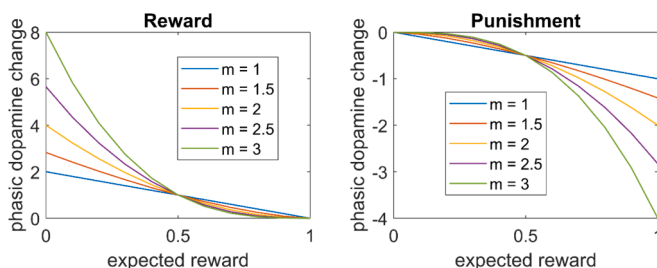


Fig. 3. Phasic changes in quantity D (Eqs. (5) and (6)) evaluated with different values of the expected reward, r_{expected} , and different values of parameter m . The curves have been obtained using $D_p = 1$. A change in D_p causes a proportional increase in the same curves.

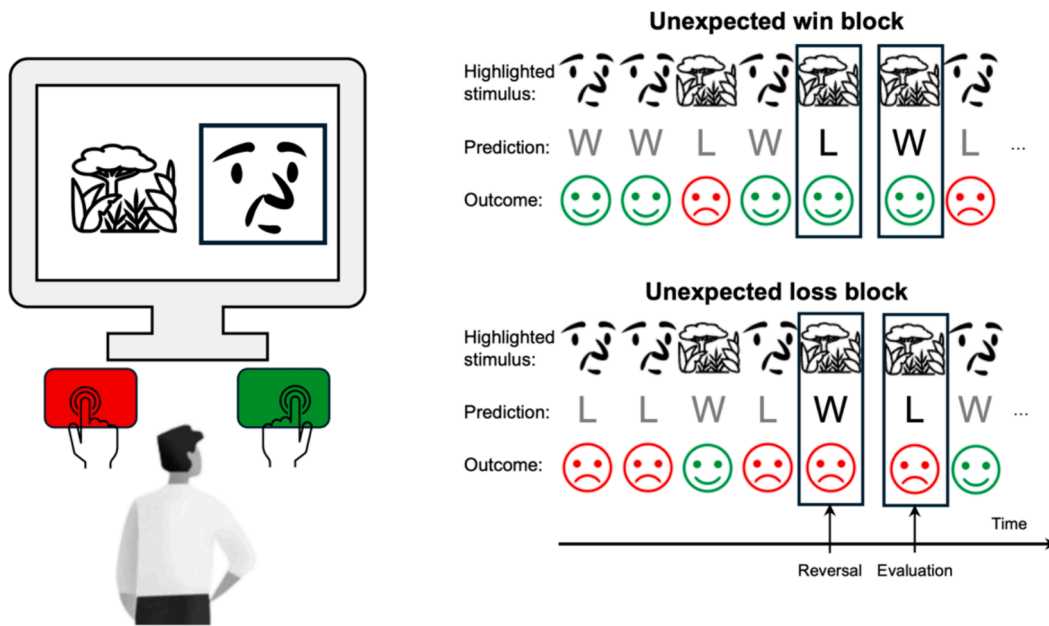


Fig. 4. Schematic representation of the task performed by Cools et al. (2006). The task consists of two separate blocks. Based on the highlighted image on the screen (hypothetical player selection), the participant should select either the green button to predict a win (W) or the red button to predict a loss (L) associated with the highlighted image. The timeline on the right shows the progression of highlighted stimuli and corresponding predictions, with “W” and “L” indicating predictions of a win and a loss, respectively, while emoticons represent the received outcomes. The first highlighted block indicates the reversal phase, where previously learned win/loss associations are switched. The second highlighted block indicates the evaluation phase, where the participant’s ability to adapt to the reversal is assessed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

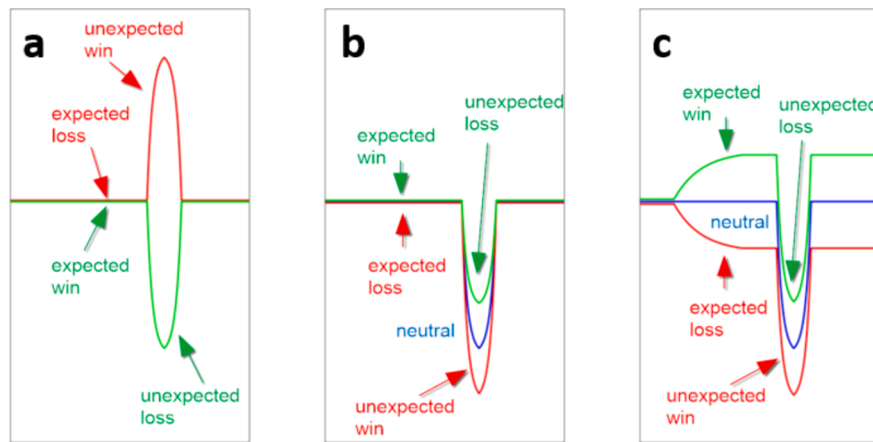


Fig. 5. Qualitative examples of the strategies adopted during the deterministic task. a assumes a single action channel, with the Go signaling the win and the NoGo signaling the loss. After an expected win (the participant pushes the green button), the unexpected loss causes a dip. After an expected loss (the participant pushes the red button), the unexpected win causes a peak. b and c assume two distinct action channels, one for wins and the other for losses. b illustrates our initial hypothesis that distinct types of prediction errors influence the phasic dips in DA. In c, the “win prediction” valence (expected win indicated by the green line) causes an increase in DA levels in the striatum. Conversely, the “loss prediction” valence (expected loss indicated by the red line) causes a decrease in DA levels, symbolizing a downward shift. In both cases (as well as in the neutral case, blue line), the unexpected achievement causes a negative peak. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

white noise because the task is deterministic; hence, low exploration should be performed by the network. In each trial, two stimuli, $S_1 = [01]$ and $S_2 = [10]$ are presented to the network. The exposure is long enough to allow the network to select a winner and subsequently reach a new steady state condition. The Hebb rule is applied at the end using steady-state values. These stimuli are randomly permuted and have a one-to-one relationship with the number of channels of the Go and NoGo pathways. After the response, the correctness of the prediction is evaluated, and phasic DA changes to the Go and NoGo neurons are computed in accordance with Eqs. 4-6. When a reward is given, it increases the activation of the winner Go neuron and decreases other Go and NoGo

neurons. At the same time, a punishment reduces the activation of all Go neurons and increases the activation of the winner NoGo neuron above the threshold. After a reversal trial, the stimulus with an unexpected outcome was always repeated in the immediately following trial. As in Cools et al. (2006), the number of trials is set to 120, and the maximum number of reversals is set to 14. The criterion used to assess the results from the network is the percentage of the errors performed by the different subjects in the trial immediately after the reversal.

2.3.2. Probabilistic task

In the experiment by Cools et al. (Cools et al., 2001), two different

colors are directly associated with rewards or punishment in a probabilistic way. In detail, selecting a color resulted in a reward 80 % of the time and a punishment 20 % of the time. In contrast, choosing the other color is rewarded with a 20 % probability and punished with an 80 % probability.

At the model level, to simulate this task, we adopted a two-choice experimental framework, i.e., the subject must choose between two possible actions: “choose the left color” or “choose the right color.” The “left color” leads to an 80 % probability of rewards, while the “right color” is associated with a 20 % probability of rewards. In this experiment, the acquisition and the reversal stages are kept distinct. In particular, after completion of the acquisition stage, which consists of 40 epochs, the reversal phase follows, in which probabilities are oppositely associated.

Translation into the model is performed by implementing one stimulus S_1 in the sensory representation set at a fixed value of 1. At the same time, in the cortex, two neurons are present, coding for the two possible actions. Similar to the deterministic task, the network starts from a totally naïve condition, wherein all synapses begin with equivalent values.

Interestingly, we used the same network parameters as in the deterministic task. The only modified parameters are the learning factor for the Hebb rule (σ in Eq. (3)) and the noise amplitude (SD) for cortical neurons (σ was reduced while SD was increased, as shown in Table I in Supplementary Materials 1). This choice is reasonable since a deterministic task requires that the new associations are learned in one shot, thus necessitating a high learning factor. In contrast, a probabilistic task requires that associations are extracted from statistics, hence using a smaller learning factor. To ensure an appropriate level of exploration by the network (hence, higher noise), Gaussian white noise with zero mean value and standard deviation (SD) of 0.15 is applied as input to the neurons of the motor cortex. The initial training phase consists of 40 trials, during which the stimulus S_1 is presented at each trial. When the network performs an action, represented by the neuron of the motor cortex overcoming a predetermined threshold of 0.9, either a reward or punishment occurs. In case of no response or multiple responses, no feedback is provided to the network. Moreover, during the training, action one is rewarded with an 80 % probability and punished with a 20 % probability; the opposite holds for action 2. After the learning stage was completed, the reversal stage followed, again consisting of 40 trials in which probabilities were oppositely associated. Results are evaluated by estimating the number of patients who successfully passed the stage; as in (Cools et al., 2001), a stage is considered successfully passed when the subject gives at least eight consecutive correct responses.

For statistical evaluation, we simulated fifty different subjects for each of the previous tasks, and each subject was characterized by a different realization of the random noise. Moreover, the tonic and/or phasic DA parameters and the parameter m were changed to realize an SA.

3. Results

3.1. Primary results on the deterministic task

First, we simulated the deterministic experiment described in Cools et al. (Cools et al., 2006). In particular, as specified in the Method section, we assessed the switch error rate, i.e., the percentage of errors in the trials immediately following an unexpected switch. It should be pointed out that we have not assigned different valences (win/loss) to the two actions during these preliminary simulations.

During the first set of simulations, some parameters of the network representing the maximum and minimum values for the synapses in the striatum w_{max} and w_{min} , respectively, the learning factor for the Hebb rule σ , the noise amplitude for cortical neurons SD were assigned to obtain percentage errors of the same order (about 10–15 %) as those reported in Cools et al. (2006). A SA on the role of these parameters,

which justifies the chosen values and their impact, can be found in Fig S4 in Supplementary Materials 2. All the other network parameters have the same value as in previous works. A value for these parameters can be found in Table I of Supplementary Materials 1.

To find appropriate values for the learning factor of the Hebb rule σ , the noise amplitude (SD) and upper and lower saturation synapses in the striatum (w_{max} and w_{min} , respectively), these parameters have been the subject of a SA (see Supplementary Materials 2). We have chosen a low value of white noise because the task is deterministic; hence, low exploration should be performed by the network. To further assess the model’s robustness to parameter changes, we performed a SA on three other parameters that affect the final choice, i.e., the strength of the winner-takes-all competition in the motor cortex, the strength of the feedback connection from the thalamus to the motor cortex, and the strength of the hyper-direct pathway. Results show that the model is quite robust: changing each of these parameters affects the percentage of errors moderately in a gradual way.

Furthermore, we performed a SA on the main factors affecting the dopaminergic response. These factors include the tonic level, D_b , the phasic coefficient in eq. 5–6, D_p , and the coefficient m in eq. 5–6, which is associated with the impact of the unexpected punishment: the higher the m , the higher the role for the unexpected punishments. Again, we remember that D does not represent DA concentration but the effect on the Go and NoGo striatal neurons.

To determine whether the literature data could be more accurately accounted for by changes in tonic D levels, by changes in the phasic D response or a combination of both, we first studied the combined effect of tonic D_t and phasic D_p parameters. Using $m = 2.5$, Fig. 6a shows the presence of a U-shaped relationship between D_t and the percentage of switch errors, consistent with previous experimental works that show a similar relationship between DA and cognition (Arnsten, 1998). However, this relationship becomes less pronounced as the phasic parameter increases (for instance, with a value of D_p equal to 1.0 in Fig. 6a). With values as high as 1.5 or 2.0, the curve becomes flat up to high values of D_t (results not shown for brevity). This suggests that increasing phasic DA changes can attenuate the effects of increasing tonic DA levels on switch errors. With a more substantial phasic DA change, the tonic DA range associated with a low fraction of wrong responses expands.

Fig. 6b shows how the percentage switch error varies with tonic DA levels, computed by maintaining the phasic coefficient $D_p = 0.8$ and using different values of m . Results indicate that the coefficient m needs to be at least equal to 2 to obtain acceptable switch errors, supporting the mechanism in eq. 4–6.

Following our SA, to simulate ON- and OFF-medicated PD patients and control subjects, we assumed that:

- i) Tonic DA level has smaller values in OFF-PD patients, intermediate values in control subjects, and elevated values in ON-PD patients (see (Cools et al., 2006) for an accurate justification);
- ii) The phasic factor D_p is substantially the same in the three groups. To maintain the U-shaped relationship, we used a value $D_p = 0.8$ since, according to Fig. 6a, it aligns with the results by Cools et al.
- iii) The parameter m is not affected by the patient’s status. We used a value $m = 2.5$ throughout the subsequent simulations, which warrants a U-shaped relationship.

The previous assumptions are summarized in the curve “switch-error” vs. tonic DA reported in Fig. 7a and used for the following tests.

3.2. The deterministic task with a different action valence

The previous results were obtained assuming no different valence (win/loss) for the two actions, i.e., only the feedback of an unexpected outcome guides the participant response. However, in the experiments by Cools et al. (2006, 2009), participants should predict whether a given card leads to winning or losing money. As discussed above, combining a

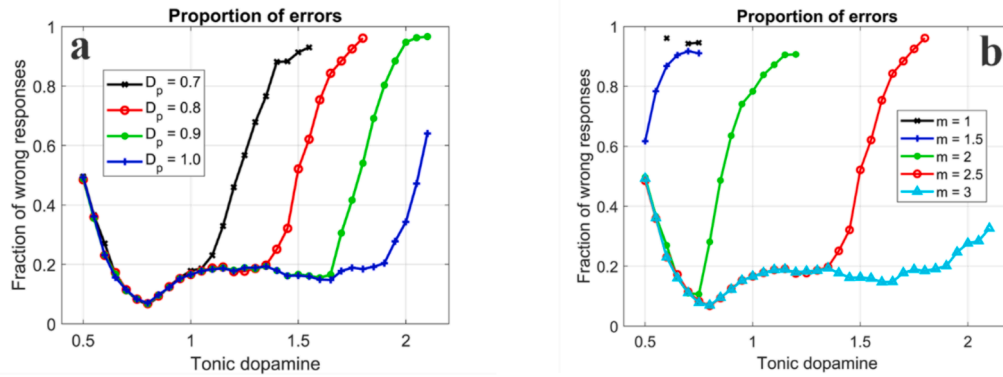


Fig. 6. Sensitivity analysis (SA) on the role of the tonic dopamine and parameters describing the phasic dopamine changes (i.e., the parameter D_p in Eqs. 5–6, a and the parameter m in Eqs. 5–6, b). The figure represents the fraction of wrong responses immediately after a switch (mean values on 50 simulated subjects) plotted as a function of tonic dopamine per different values of parameters D_p and m . A U-shape relationship is evident for values of D_p less than one and values of m greater than 2. In the case of m as low as 1—1.5 and high tonic dopamine levels, many wrong responses occur (hence, the curves are not plotted in this range).

win/loss and positive/negative feedback can be interpreted as able to reinforce either the Go or the NoGo of a single action separately (i.e., the Go would signal the win, and the NoGo would signal the loss, see Fig. 5a). Alternatively, we assume that prediction of reward/punishment and instrumental control of action are signaled by two distinct Go and two distinct NoGo channels.

In a series of papers, Cools et al. observed that PD-ON subjects made fewer errors after an unpredicted win (i.e., when a previously losing card now wins) as compared to more errors when switching after an unpredicted loss (i.e., when a previously winning card now becomes a loser). The opposite error pattern, although less evident, is observable in PD-OFF patients, who showed fewer errors after an unexpected loss than after an unexpected win. The results by Cools et al. (2006) are summarized in Fig. 7b.

In order to explain these results, we first assumed that phasic DA dips differ according to valence, with a deeper dip after an unexpected win and a shallower one after an unexpected loss, compared to the neutral condition (Fig. 5b). As it can be seen by looking at Fig. 6a, this difference in phasic response according to valence can explain the better adaptation of ON-medicated patients (higher values of D_p) to unexpected wins than unexpected loss; however, it cannot explain the opposite behavior observed in OFF-medicated subjects (smaller values of D_p , where the curves for different values of D_p are overlapped). Furthermore, minimal differences in parameter D_t are sufficient to cause enormous differences between control and ON-medicated patients, which seems quite unrealistic. To show an example, we assumed a change of parameter D_p from 0.85 to 0.8 for the two valences and values of $D_t = [0.9 \ 1.35 \ 1.45]$ for OFF-medicated patients, Control subjects, and ON-medicated patients, respectively. The results can be found in Fig. S5 of Supplementary Materials 2.

Different authors in recent years (Dayan and Huys, 2008; Niv et al., 2007; Rigoli et al., 2016a) suggested that the tonic DA level, or more generally the basal condition of the BG, can be actively regulated to signal the valence of the response; particularly, higher DA levels would promote cognitive effort (Westbrook et al., 2021).

Based on these ideas, building upon our SA and the U-shaped relationship (Fig. 7a), we introduced a different assumption, i.e., the valence of the stimulus can moderately affect the parameter D_t in anticipation of the response, setting the basal value at a slightly higher or lower level just before the stimulus (see Fig. 5c). As shown in Fig. 7a, we assume that, if a stimulus signals an *expected reward*, the value of this parameter moves to a slightly higher value (we can use $D_t + v$, where v signals a positive valence of the stimulus). During a reversal, this condition is followed by an *unexpected punishment*. Conversely, if a stimulus signals an *expected punishment*, the value of parameter D_t moves to a slightly smaller value (we can say $D_t - v$, i.e., a negative valence of the stimulus).

During reversal, this condition is followed by an *unexpected reward*. This allows for the simulation of three groups (PD-OFF, PD-ON, and Control). By using an interval of values for D_t along the curve, with different upper and lower values for each group, we tested if we were able to reproduce the results observed by Cools et al. (2006) using our model.

To implement this idea, we assumed values for the parameter D_p , equal to 0.65, 1.0, and 1.4 for OFF-medicated PD patients, control subjects, and ON-medicated PD subjects, respectively. To mimic the two valence conditions of unexpected punishment and reward, these values were incremented and decremented by 0.05, respectively (i.e., using $v = 0.05$). The results, reported in Fig. 7c (and further put in evidence in the U-shaped curve of Fig. 7a), agree with the results by Cools et al. quite well not only for what concerns differences of the response in ON-PD and OFF-PD subjects but also in control individuals.

Finally, Fig. 7d displays data from (Cools et al., 2009) showing the difference in healthy subjects between the proportion of correct responses in switch trials after an unexpected reward and those after an unexpected punishment as a function of their relative DA synthesis. Subjects with higher DA synthesis exhibit better reversal learning after unexpected rewards, whereas subjects with low DA synthesis show the opposite pattern. As reported in Fig. 7e, a similar quasi-linear relationship can be obtained from our model as a function of the D_b , simply making use of the previous assumption, i.e., we used $D_t + 0.05$ in case of unexpected losses and $D_t - 0.05$ in case of unexpected wins (where D_t is the value plotted in the abscissa of Fig. 7e). A change in parameter D_p (phasic DA) cannot explain these results, producing significant changes for high D_b , but insignificant changes for low D_b .

3.3. Simulation of the probabilistic reversal learning

In a second set of simulations, we replicated the probabilistic reversal learning task described in (Cools et al., 2001). Here, we maintained the same network parameters as in the deterministic task, modifying only the learning factor for the Hebb rule and the noise amplitude for cortical neurons, as required by the nature of the task.

Fig. 8a and 8b report a SA on the tonic D_t level and the phasic factor ($D_p = 0.7$ and $D_p = 0.8$, respectively). The values $D_p = 0.9$ and $D_p = 1.0$ were also tested, producing almost flat curves for the reversal case, i.e., with more than 90 % of people passing. Results are expressed as the number of subjects who passed the acquisition task and the reversal task (Cools et al., 2001). These results show that a progressive increase in tonic D_t is associated with a significant inability to perform the reversal task. Finally, an SA on the parameter m is shown in Fig. 8c (with $D_p = 0.7$ and $D_t = 1.2$), confirming that values of m as high as 2.5 are necessary to realize an accurate reversal.

Fig. 8d shows the results by (Cools et al., 2001) involving OFF-

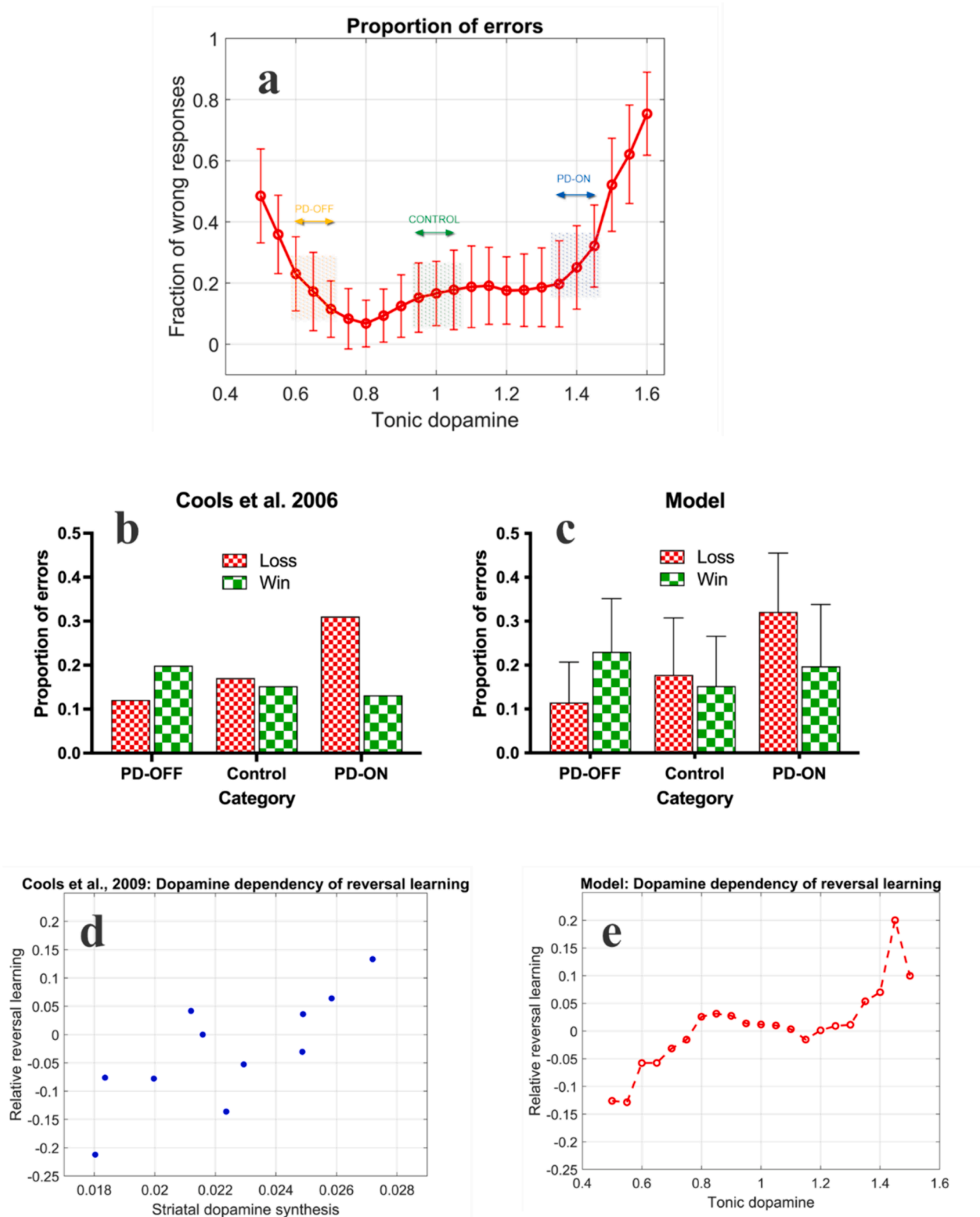


Fig. 7. Simulation of the experimental results by Cools et al. (2006 and 2009). a describes the model relationship between tonic dopamine and the fraction of wrong responses after a switch (mean values and SD, computed on 50 simulated subjects) using $D_p = 0.8$ and $m = 2.5$. In the figure, the tonic dopamine values used to simulate OFF-medicated PD patients, control subjects, and ON-medicated PD patients are also marked (the upper values hold for unpredicted punishments, the lower values for unpredicted rewards). b and c compare the results obtained by Cools et al. (2006) on PD-OFF medicated patients, control subjects, and PD-ON medicated patients and by the model (2009) using the values marked in a for the three cases (the higher tonic dopamine is used for unpredicted punishments and the lower tonic dopamine for unpredicted rewards). d and e compare the results obtained by Cools et al. (2009) (2009) (6d) as a function of dopamine production rate with the model predictions (6e). The relative reversal learning scores represent the proportion of correct responses on switch trials after unexpected reward minus the proportion of correct responses on switch trials after unexpected punishment. The model assumption was that parameter D_t was 0.1 higher in the case of unexpected punishments than in unexpected rewards.

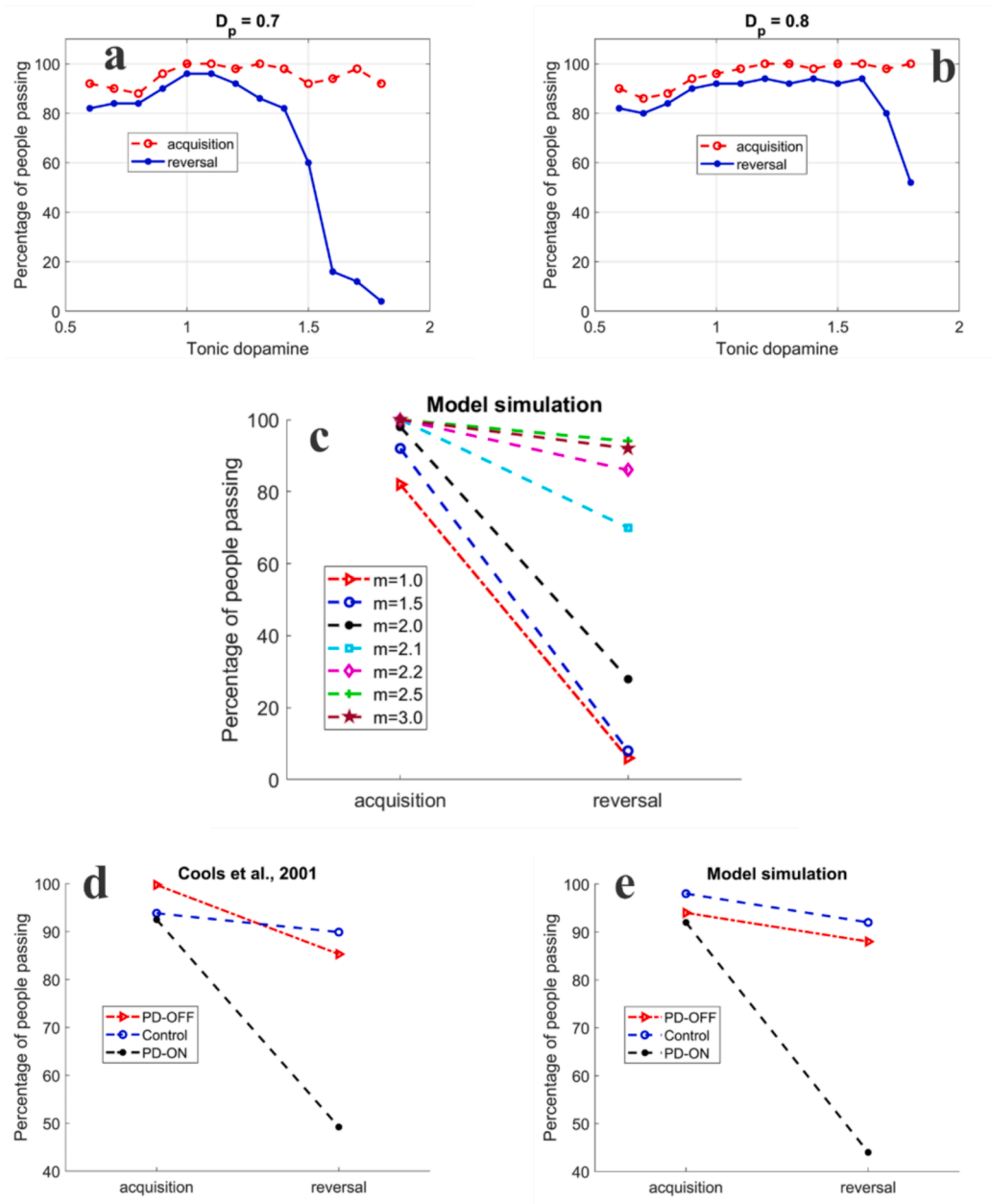


Fig. 8. Simulation of the experimental results by Cools et al. (2001). Fig. 8a and 8b represent the percentage of subjects passing the task (computed with the model on 50 simulated subjects) during the acquisition and reversal phases, plotted vs. the basic dopamine level with two different values of parameter D_p . The value of m was set at 2.5. c represents the number of subjects passing the task during the acquisition and reversal phases per different values of the parameter m . The tonic dopamine was set at $D_t = 1.2$, and the phasic parameter $D_p = 0.7$. d and e compare the results obtained by Cools et al. (2001) on PD-OFF medicated patients, control subjects, and PD-ON medicated patients (7d) with the values obtained with the model ($D_p = 0.7$; $m = 2.5$; $D_t = [0.85 \ 1.2 \ 1.52]$ for the three cases).

medicated PD patients, ON-medicated PD patients, and Control subjects, expressed again as the number of subjects who passed the acquisition task and the reversal task. Based on the SA, to mirror these patterns, we assumed a phasic factor $D_p = 0.7$, and values for D_t for the three classes (PD-OFF, PD-ON, and Controls) equal to 0.85, 1.2, and 1.52, respectively. The results of our simulations, summarized in Fig. 8e, correspond well with the data from (Cools et al., 2001), showing a similar trend in the acquisition and reversal phases across the three groups, specifically the impairment in probabilistic reversal learning in PD-ON.

4. Discussion

Neurocomputational models of the BG have a long tradition, providing significant insight into the role of the BG in action selection,

whether in physiological or pathological conditions (Baston and Ursino, 2015; Cutsuridis and Perantonis, 2006; Frank, 2006, 2005; Gurney et al., 2001a, 2001b; Humphries et al., 2018; Kato and Morita, 2016; Liu et al., 2018; Morita and Kato, 2018; Moustafa et al., 2014; Moustafa and Gluck, 2011; O'Reilly and Frank, 2006; Schroll and Hamker, 2013; Suryanarayana et al., 2019). Most of these models share some essential points, including the use of segregated channels for individual actions subdivided between a Go and NoGo pathway, Hebbian potentiation and depotentiation mechanisms for the synapses in the striatum, and a role of tonic and phasic DA changes during learning. The latter aspect of DA inclusion gave important indications into differences between control and Parkinsonian subjects (Baston and Ursino, 2015; Guthrie et al., 2009; Moustafa and Gluck, 2011; Véronneau-Veilleux et al., 2021), medicated and non-medicated PD (Frank et al., 2004; Moustafa et al.,

2014), as well as stable and unstable levodopa responders (Baston et al., 2016; Ursino et al., 2020a). Several models also provided some stimulating hypotheses on aberrant learning in PD (Beeler et al., 2013; Ursino and Baston, 2018), possible factors affecting dyskinesia (Ursino et al., 2020b), beta or gamma oscillations in PD (Humphries et al., 2006; Liu et al., 2018) and the role of tonic DA in exploration–exploitation control (Humphries et al., 2012).

The present study follows the previous neurocomputational BG tradition but focuses on a few challenges, which, in our opinion, still deserve clarification and accurate analysis. These mainly concern the way subjects can rapidly reverse their choice after a punishment (or after the absence of an expected reward), the role of phasic and tonic DA changes in learning, how expectation can be coded in the BG, and the valence of performed choices.

These aspects, along with the Hebbian mechanisms in the striatum, Go, and the NoGo pathways' role, learning rate, and noise in a deterministic and probabilistic scenario are discussed below.

Role of learning rate and noise – Interestingly, we simulated the deterministic and probabilistic reversal tasks with the same model parameters but with a moderate change in the tonic and phasic DA terms and a significant change in noise and learning rate. The first aspect can be justified by individual variability. More interestingly, different values for the noise and learning rate set an exploration/exploitation trade-off and are essential to distinguish a deterministic task from a probabilistic one. As well known, noise in cortical neurons can be used to represent the exploration capacity of a subject. Low noise indicates that a subject makes strong use of current synapse values (exploitation) but implements a reduced exploration of alternative choices. Still, high noise indicates that the subject considerably explores alternative possibilities beyond past knowledge. A deterministic task requires low noise since just one punishment should ultimately drive future behavior. Probabilistic tasks require higher noise since a continuous exploration of different alternatives is a prerequisite to learning correct statistics. For the same reasons, a high learning rate in the deterministic task sets an extreme exploitation, and a lower learning rate in the probabilistic task sets a minor exploitation vs. exploration.

Future work will determine how subjects set these values. We claim that the prefrontal cortex plays a pivotal role in this setting process. However, additional research is needed to analyze this aspect. Furthermore, we should consider the influence of neurotransmitters on these learning parameters. Specifically, the regulation of noise might be linked to the norepinephrine system.

Computation of the expectancy – A fundamental problem in reinforcement learning is how a network computes the expectancy (expected reward or expected punishment) based on previous experience. While some studies compute the expected value using an algorithm external to the model (a critic module, separated from the BG), in the present and past studies, we propose a different original approach. Reward expectancy can be evaluated as the activity (normalized between 0 and 1) of the winner Go neuron when a response is established (i.e., the moment when the winner neuron in the cortex overcomes a threshold, signaling that a choice has been made). Subsequently, this value is exploited in Eqs. (5) and (6) to drive a phasic peak or phasic dip for the DA effect on the striatum. While the expected reward in our model is wholly computed within the BG and exploits past knowledge stored in the synapses, the phasic changes simulate the response of dopaminergic neurons in the substantia nigra.

An essential parameter in our computation of phasic DA changes is the exponential m in Eqs. (5) and (6). The higher this value, the stronger the effect of an unexpected reward (or an unexpected punishment). Our SA suggests that, to simulate reversal learning correctly, this parameter must be set to values higher than 2, confirming that a strongly non-linear relationship between expectancy and phasic DA change is required.

Simulation of ON-medicated and OFF-medicated PD patients: role of tonic and phasic dopamine– An area not yet clearly defined in the current research is the role of tonic and phasic DA in PD. It is commonly known

that PD OFF-medicated patients typically exhibit lower levels of tonic DA. Conversely, based on the DA overdose hypothesis (Cools et al., 2022), PD patients ON-medicated may exhibit elevated tonic DA levels in the ventral striatum. However, the impact of PD on phasic DA changes is still unclear. Do changes in tonic DA levels in PD correspond to similar changes in phasic DA levels? This direct correlation was the assumption of our previous works (Baston and Ursino, 2015; Schirru et al., 2022). However, our latest SA suggests that to explain the U-shaped relationship between reversal errors and tonic DA observed experimentally (Arnsten, 1998; Cools et al., 2006), it seems necessary to consider that phasic DA changes are not positively correlated with tonic levels. Further, there is evidence suggesting an inverse relationship between tonic and phasic DA levels (Grace, 2016, 2001, 1991). This is particularly clear in cases of high tonic DA, such as in ON-medicated PD patients. Our SA shows that high phasic DA levels lead to a flat curve in Fig. 6a. This indicates that an increased phasic dip could potentially offset the effects of tonic DA overdose, theoretically enabling ON-medicated PD patients to respond to punishments similar to control subjects. However, this contradicts the results of several experimental studies (Bódi et al., 2009; Cools et al., 2009, 2001; Frank et al., 2004). Consequently, we can conclude that high tonic DA levels do not lead to a more pronounced negative phasic response, suggesting a relative independence of the phasic response from the tonic level.

The valence of the response – The relationship between the valence of the action choice and the quantity D_t is a new aspect of this study that still requires further validation.

In an attempt to explain the differences between losses and wins in Cools et al. (Cools et al., 2009, 2006; Robinson et al., 2010), this study proposes a preliminary but stimulating new hypothesis. As discussed earlier in the method section, there are two possible interpretations for the results of these tasks, in which it is essential to consider that the task participant is distinct from a player experiencing wins and losses:

- i) the participant uses just one action channel to predict the player's wins and losses: correct predictions of the wins reinforce the Go pathway, and correct predictions of the losses reinforce the NoGo pathway. The opposite synaptic changes occur in case of incorrect predictions. According to this schema, OFF-medicated patients have difficulty updating the Go pathway; ON-medicated patients have difficulty updating the NoGo one. This assumption is implicit in most previous papers (see Robinson et al., 2010).
- ii) Starting from a hypothetical neutral task, in which only colors are associated with a prediction without a positive or negative valence, we suggest a different interpretation. The task participant uses two segregated channels (one for the red choice and one for the green choice) with a winner-takes-all competition in the cortex. In case there is no valence for the two choices, the two channels should be symmetrical, behaving similarly. Differences between the two choices (green = wins; red = losses) should depend on a valence-dependent response bias, which makes the two channels asymmetrical.

By comparing the model and experimental results (Cools et al., 2009, 2006), we thought about two possible ways to introduce this response bias, acting on the quantity D in the model. One possibility may be to assume that the phasic parameter D_p is different, being higher after an unexpected reward and smaller after an unexpected loss. However, this choice cannot explain the difference observed in OFF-medicated patients (see Fig. 6a and Fig. S5 in Supplementary Materials 2). Looking at Fig. 7a, we propose that the valence bias in the response can be explained by adding a valence contribution to the parameter D_t in the model (that is, using $D_t \pm v$ in the simulations). In particular, as shown in Fig. 5c, an expected win (which, during a switch, is followed by an unexpected loss) is associated with a positive valence ($v = 0.05$). In contrast, an expected loss (followed during the switch by an unexpected win) is associated with a negative valence ($v = -0.05$).

A question is: what may be the origin of this “valence” quantity v ? We remember that the quantity D in our model represents an *effect* acting on the Go and NoGo striatal neurons. One possibility is that v actually represents a change in DA concentration. In past years, various studies suggested that tonic DA level signals the valence of an action. In particular, Niv et al. (2007) used psychological and computational methods to establish a link between higher levels of DA with a more vigorous response. Zénon et al. (2016) proposed that DA is related to the effort necessary to reach a given goal. Rigoli et al. (2016b) reported that boosting DA levels increases the propensity for gambling and the attractiveness of risky actions. Rigoli et al. (2016a) observed that the prospect of punishment is typically characterized by below baseline levels of dopaminergic function and found that neural responses in the ventral striatum and ventral tegmental area/substantia nigra covaried with the expected value. Saunders et al. (2018) observed that, after training, conditioned stimuli can evoke DA neuron activity on their own, thus instantiating a motivational signal. Niv et al. (2007) highlights the role of a tonic signal in determining the optimal rate of responding, implying a tight coupling between motivational states and tonic DA. However, it is also possible that this valence signal is related to serotonin (Dayan and Huys, 2008) or norepinephrine. The innervation of the BG by the serotonin system is discussed in Parent et al. (2011).

More generally, any signal that affects the basal condition of the Go and NoGo pathways can be represented by our variable v in the model. In fact, it is important to stress that we are assuming the existence of two different signals: one, related to phasic changes, ΔD , which should implement predictions (hence, a negative dip after any wrong prediction), and another one (v affecting D) related with the valence (hence wins or losses). Interestingly, Robinson et al. (2010), examining both responses in a similar paradigm, observed two distinct signals: “In addition to the prediction mechanism, subjects might have additionally recruited an instrumental mechanism, which is likely driven primarily by a positive, reward-signed signal associated with the state in which punishment is not expected.”

Previous studies on dopaminergic modulation and reversal learning – Several recent studies have analyzed DA modulation during reversal learning. The main conclusions qualitatively agree with a few assumptions of the present work and can be summarized through the following main points: i) DA release simultaneously encodes cost, benefit, and motivation (Eshel et al., 2024) and interacts with a network of cortical regions, representing not only reward but also a more complex strategy (Calabro et al., 2023); ii) DA exhibits both transient kinetics and slowly developing signals (Salinas et al., 2023), with motivation for rewards reflecting a state that changes over slower timescales (Eshel et al., 2024) with a role for tonic DA (Delaney et al., 2024; Wang et al., 2021). This point partially supports the basic idea shown in Fig. 5c; iii) D1 and D2 receptors have complementary functions in learning (Kwak and Jung, 2019; Sala-Bayo et al., 2020; Verharen et al., 2019). In particular, D2 receptor damage impairs reversal learning by blocking the impact of negative feedback (Alsiö et al., 2019; Kruzich et al., 2006). The latter results confirm what was suggested in Fig. 2, where a clear role for the NoGo during inhibition is evident.

Comparison with previous modeling papers – Although several BG neurocomputational models have appeared in recent years, only a few of them were explicitly related to reversal learning. The probabilistic reversal learning by Cools et al. (2001) has been simulated by Moustafa et al. (2014), reporting almost similar results to our Fig. 8. However, in that model the subjects’ disease status and dopaminergic medications were modulated by means of four parameters: two learning rates (one for the BG and the other for the PFC) and two gain parameters (BG and PFC modules) without an explicit description of tonic and phasic DA terms. Frank (2005) later explicitly simulated DA within their model. Their simulation included the overdose case, showing impaired probabilistic reversal. However, to our knowledge, there are no models that simulate both probabilistic and deterministic reversal tasks using the same model framework and a single set of parameters.

Furthermore, our model shares similarities with the one developed by Humphries et al. (2012). These authors proposed that the outputs of the BG could be interpreted as a probability distribution function for action selection. Here, we adopted a similar concept by using the Winner Go activity as an indicator of reward expectation. Moreover, Humphries et al. (2012) suggested that tonic striatal DA influences the exploration–exploitation trade-off, with increased tonic striatal DA reducing the level of exploration. In our model, a task-dependent difference in exploration–exploitation is established using different values of noise and learning rate, while quantity D , sets a “valence dependent” difference. However, the two approaches are pretty similar.

In the model by Guthrie et al., (2013), two distinct but interacting loops for cognitive and motor functions are created. They incorporate mechanisms like synaptic noise and DA-modulated learning to model how decisions in one part of the brain (cognitive) can influence decisions in another part (motor). This approach helps to understand how complex decision-making processes are coordinated in the brain. This model shares similarities with our model in the way the phasic DA signals are used to adjust the synaptic weights. However, there are differences in the specific task each model simulates. Van Swieten and Bogacz (2020) explore the effect of motivation on choice and learning, integrating reinforcement learning with incentive salience theory to explain how physiological states like hunger influence action selection. This may align with our consideration of dopaminergic modulation and state-dependent learning.

Moreover, Mikhael et al. (2022) use a BG model to investigate state uncertainty’s influence on DA dynamics, showing that sensory feedback causes DA reward prediction errors to ramp up. The results are also supported by their theoretical predictions with empirical work in mice. While not directly focused on reversal tasks, this aligns with our emphasis on the role of DA signal during reward expectations and our model’s anticipatory signal mechanism governed by tonic DA. Maith et al., (2023) investigated the role of STN/GPe synaptic plasticity in exploration behavior after reversals, suggesting the involvement of multiple regions within the BG and reduced independence of the three main pathways. A new synaptic plasticity rule showed that exploration becomes biased towards previously rewarded positions. Their task, involving a more complex 5-choice reversal learning paradigm, emphasizes the complexity of exploratory behavior. They derived a learning rule for STN/GPe connections to accumulate prior experience via synaptic plasticity, whereas our model trains the synapses in the striatum.

Unique model characteristics – Despite the presence of many previous models, we think the current work introduces some unique contributions, summarized below:

- (i) to our knowledge, this is the only model that simulates learning during both deterministic and probabilistic tasks with a single parameter set;
- (ii) the model introduces the strong assumption that tonic DA represents an anticipatory signal for the valence of choice (positive or negative), which is different from the participant reward expectation;
- (iii) we suggest that a complete Hebb rule, similar to that used in classic auto-associative networks, can explain reversal learning on the striatum during different tasks.

Testable predictions and experimental validation – Starting from the previous considerations, we suggest new tasks (or variants of already proposed tasks) to disentangle the hypothesis that wins and losses are related to GO and NoGO in the same action channel (hence DA peaks and dips, as in the interpretation of Cools et al.) from the present hypothesis involving a role for action valence in two separate channels. The idea is to perform a two-choice task, associated with two sensory images, in which the participant must predict if a given episode will occur (response YES or green button) or will not occur (response NO or red button), depending on the presented image. For instance, the images can be the faces of two individuals, and the participant must predict which individual will experience the episode. During each task, reversals

can be performed multiple times to evaluate switch errors. The same task can be performed in control subjects and PD-ON and PD-OFF patients. Differences in the percentage of errors and latency in the response can be compared with model predictions. The same task can be repeated three times with three different valences of the occurring episode to validate the present hypothesis.

Task i) Neutral valence: the participant predicts if the individual in the image will perform a neutral action (e.g., wearing a hat or not). Task ii) Negative valence: the participant should guess whether the individual presented in the image will experience a negative event (e.g., having a severe accident). In this condition, the model assumes a decrease in tonic DA during the yes prediction, leading to improved reversal responses in PD-ON patients compared with the previous case, and worsened responses in PD-OFF patients. Task iii) Positive valence: the participant predicts whether the individual in the image will experience a positive event (e.g., being happy). In this condition, the model assumes an increase in DA during the yes prediction, leading to more switch errors in PD-ON patients. The previous tasks can be either deterministic, with many switches after a random number of correct responses (as in Cools et al., 2006), or probabilistic, with a single switch after a sufficient number of trials. Furthermore, if neuroimage data on striatal DA synthesis capacity were available, the accuracy in the same tasks could be assessed against the DA synthesis level.

Model limitations and future research – The model exhibits some limitations, which can be the target of future studies. First, it does not distinguish between the dorsal and ventral parts of the striatum. In PD patients, the DA level is probably more severely reduced in the dorsal part than in the ventral. Hence, during levodopa treatment, DA overload probably holds for the ventral portion only, implicated in cognitive decisions, whereas the dorsal portion, more implicated in motor responses, is less affected. Future models could address this by developing two distinct BG models to simulate varying levels of denervation in the dorsal and ventral striatum. This can help to investigate the hypothesis that different parts of the BG are implicated in motor response control and motivational control, corresponding to loops involving the sensorimotor cortex and the dorsal striatum and those involving the frontal/limbic cortex and the ventral striatum, respectively. Such models would help examine the differential impacts of dopaminergic dysfunction in these two striatal regions, offering more profound insights into PD pathology and its influence on cognitive and motor functions.

Furthermore, the present model uses a simplified description of the DA effect on the Go and NoGo parts. Future versions may incorporate a detailed description of the DA release in the Substantia Nigra and D1 and D2 receptor characteristics in the striatum. This may allow a deeper understanding of the changes in tonic and phasic DA and their relative relationships and a more accurate analysis of drug effects. The present model also does not include a direct control by the orbitofrontal cortex, which seems to incorporate the relative motivational significance of different rewards (Hollerman et al., 2000). Finally, a more sophisticated synaptic dynamics could be used, to explain the presence of beta oscillations in the BG, which are implicated in normal movement suppression and motor impairment in PD.

Recent studies such as those by Isoda and Hikosaka (Isoda & Hikosaka, 2008) showed that STN neurons are involved in both stopping and facilitating responses, suggesting a more complex bidirectional circuit with the GPe. In addition, a study (Wang et al., 2019) demonstrated that optogenetic stimulation of striatal neurons in the indirect pathway during reversal learning increases thalamic activity, challenging the classical view of the indirect pathway's role. These findings can be studied further within the model to explore the implications of these complex pathways.

Future research directions also include using the model to simulate additional pathological conditions involving the BG and the dopaminergic system, such as ADHD, Huntington's disease, or schizophrenia. Finally, it is well known that the BG form connections and circuits with other brain regions (such as the prefrontal cortex and the cerebellum),

and these interactions are crucial in several motor, cognitive, and affective functions. Specifically, a new functional perspective is that the BG, the cerebellum, and the cerebral cortex form an integrated network (Bostan and Strick, 2018). Future extensions of the model may involve interconnections between the BG and the prefrontal cortex and between the BG and the cerebellum, or even creating of an integrated network among all these regions working together.

Funding source Work supported by #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006) – A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022).

Also supported by the project “The effect of emotions on associative memory in Parkinson disease: from behavioral to computational approach” – code 2022LLCH97 – founded by the European Union – NextGenerationEU National Recovery and Resilience Plan (NRRP) – PRIN 2022.

MS is grateful for the financial support provided by the Fonds de recherche du Québec – Nature et technologies (FRQNT) and from the Centre Interdisciplinaire de Recherche sur le Cerveau et l'Apprentissage (CIRCA).

Author contributions

Conceptualization: MU, DA, SP, MS. Funding acquisition: MU, SP, MS; Methodology: MU, SP, FN, PR, MS; Software: SP, MS; Supervision: MU, FN, PR. Validation: MU, SP. Visualization: MU, SP, MS. Writing – original draft: MU, SP. Writing – review & editing: MU, SP, FN, PR, MS.

CRediT authorship contribution statement

Mauro Ursino: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Silvana Pelle:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Investigation, Data curation. **Fahima Nekka:** Writing – review & editing, Supervision, Methodology. **Philippe Robaey:** Writing – review & editing, Supervision, Methodology. **Miriam Schirru:** Writing – review & editing, Visualization, Software, Methodology, Investigation, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.nlm.2024.107985>.

References

- Alsö, J., Phillips, B. U., Sala-Bayo, J., Nilsson, S. R. O., Calafat-Pla, T. C., Rizwand, A., Plumbridge, J. M., López-Cruz, L., Dalley, J. W., Cardinal, R. N., Mar, A. C., & Robbins, T. W. (2019). Dopamine D2-like receptor stimulation blocks negative feedback in visual and spatial reversal learning in the rat: Behavioural and computational evidence. *Psychopharmacology (Berl)*, 236, 2307–2323. <https://doi.org/10.1007/s00213-019-05296-y>
- Arnsten, A. F. T. (1998). Catecholamine modulation of prefrontal cortical cognitive function. *Trends in Cognitive Sciences*, 2, 436–447. [https://doi.org/10.1016/S1364-6613\(98\)01240-6](https://doi.org/10.1016/S1364-6613(98)01240-6)
- Boston, C., Contin, M., Calandra Buonaura, G., Cortelli, P., & Ursino, M. (2016). A Mathematical Model of Levodopa Medication Effect on Basal Ganglia in

- Parkinson's Disease: An Application to the Alternate Finger Tapping Task. *Front. Hum. Neurosci.*, 10, 280. <https://doi.org/10.3389/fnhum.2016.00280>
- Baston, C., & Ursino, M. (2015). A Biologically Inspired Computational Model of Basal Ganglia in Action Selection. *Comput. Intell. Neurosci.*, 2015, Article 187417. <https://doi.org/10.1155/2015/187417>
- Beeler, J. A., Petzinger, G., & Jakowec, M. W. (2013). The Enemy within: Propagation of Aberrant Corticostriatal Learning to Cortical Function in Parkinson's Disease. *Front. Neurol.*, 4, 134. <https://doi.org/10.3389/fneur.2013.00134>
- Bódi, N., Kéri, S., Nagy, H., Moustafa, A., Myers, C. E., Daw, N., Dibó, G., Takáts, A., Bereczki, D., & Gluck, M. A. (2009). Reward-learning and the novelty-seeking personality: A between- and within-subjects study of the effects of dopamine agonists on young Parkinson's patients*. *Brain*, 132, 2385–2395. <https://doi.org/10.1093/brain/awp094>
- Bostan, A. C., & Strick, P. L. (2018). The basal ganglia and the cerebellum: Nodes in an integrated network. *Nat Rev Neurosci.*, 19, 338–350. <https://doi.org/10.1038/s41583-018-0002-7>
- Calabresi, P., Picconi, B., Tozzi, A., Ghiglieri, V., & Di Filippo, M. (2014). Direct and indirect pathways of basal ganglia: A critical reappraisal. *Nat Neurosci.*, 17, 1022–1030. <https://doi.org/10.1038/nn.3743>
- Calabro, F. J., Montez, D. F., Larsen, B., Laymon, C. M., Foran, W., Hallquist, M. N., Price, J. C., & Luna, B. (2023). Striatal dopamine supports reward expectation and learning: A simultaneous PET/fMRI study. *Neuroimage*, 267, Article 119831. <https://doi.org/10.1016/j.neuroimage.2022.119831>
- Cohen, M. X., & Frank, M. J. (2009). Neurocomputational models of basal ganglia function in learning, memory and choice. *Behav. Brain Res.*, 199, 141–156. <https://doi.org/10.1016/j.bbr.2008.09.029>
- Cools, R., Altamirano, L., & D'Esposito, M. (2006). Reversal learning in Parkinson's disease depends on medication status and outcome valence. *Neuropsychologia*, 44, 1663–1673. <https://doi.org/10.1016/j.neuropsychologia.2006.03.030>
- Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2001). Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. *Cereb Cortex*, 11, 1136–1143. <https://doi.org/10.1093/cercor/11.12.1136>
- Cools, R., Frank, M. J., Gibbs, S. E., Miyakawa, A., Jagust, W., & D'Esposito, M. (2009). Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. *J Neurosci.*, 29, 1538–1543. <https://doi.org/10.1523/JNEUROSCI.4467-08.2009>
- Cools, R., Tichelaar, J. G., Helmich, R. C. G., Bloem, B. R., Esselink, R. A. J., Smulders, K., & Timmer, M. H. M. (2022). Role of dopamine and clinical heterogeneity in cognitive dysfunction in Parkinson's disease. *Prog Brain Res.*, 269, 309–343. <https://doi.org/10.1016/bs.pbr.2022.01.012>
- Coulthard, E. J., Bogacz, R., Javed, S., Mooney, L. K., Murphy, G., Keeley, S., & Whone, A. L. (2012). Distinct roles of dopamine and subthalamic nucleus in learning and probabilistic decision making. *Brain*, 135, 3721–3734. <https://doi.org/10.1093/brain/awt273>
- Cutsuridis, V., & Perantonis, S. (2006). A neural network model of Parkinson's disease bradykinesia. *Neural Netw.*, 19, 354–374. <https://doi.org/10.1016/j.neunet.2005.08.016>
- Dagher, A., & Robbins, T. W. (2009). Personality, Addiction, Dopamine: Insights from Parkinson's Disease. *Neuron*, 61, 502–510. <https://doi.org/10.1016/j.neuron.2009.01.031>
- Dayan, P., & Huys, Q. J. M. (2008). Serotonin, inhibition, and negative mood. *PLoS Comput Biol.*, 4, e4.
- Delaney, J., Nathani, S., Tan, V., Chavez, C., Orr, A., Paek, J., Faraji, M., Setlow, B., & Urs, N. M. (2024). Enhanced cognitive flexibility and phasic striatal dopamine dynamics in a mouse model of low striatal tonic dopamine. *Neuropsychopharmacology*. <https://doi.org/10.1038/s41386-024-01868-5>
- Delignat-Lavaud, B., Kano, J., Ducrot, C., Massé, I., Mukherjee, S., Giguère, N., Moquin, L., Lévesque, C., Burke, S., Denis, R., Bourque, M.-J., Tchung, A., Rosaneto, P., Lévesque, D., De Beaumont, L., & Trudeau, L.-É. (2023). Synaptotagmin-1-dependent phasic axonal dopamine release is dispensable for basic motor behaviors in mice. *Nat Commun.*, 14, 4120. <https://doi.org/10.1038/s41467-023-39805-7>
- Di Filippo, M., Picconi, B., Tantucci, M., Ghiglieri, V., Bagetta, V., Sgobio, C., Tozzi, A., Parnetti, L., & Calabresi, P. (2009). Short-term and long-term plasticity at corticostriatal synapses: Implications for learning and memory. *Behav Brain Res.*, 199, 108–118. <https://doi.org/10.1016/j.bbr.2008.09.025>
- Diederer, K. M. J., Ziauddeen, H., Vestergaard, M. D., Spencer, T., Schultz, W., & Fletcher, P. C. (2017). Dopamine Modulates Adaptive Prediction Error Coding in the Human Midbrain and Striatum. *J Neurosci.*, 37, 1708–1720. <https://doi.org/10.1523/JNEUROSCI.1979-16.2016>
- Driver-Dunckley, E., Samanta, J., & Stacy, M. (2003). Pathological gambling associated with dopamine agonist therapy in Parkinson's disease. *Neurology*, 61, 422–423. <https://doi.org/10.1212/01.WNL.0000076478.45005.EC>
- Eshel, N., Touponce, G. C., Wang, A. R., Osterman, A. K., Shank, A. N., Groome, A. M., Taniguchi, L., Cardozo Pinto, D. F., Tucciaron, J., Bentzley, B. S., & Malenka, R. C. (2024). Striatal dopamine integrates cost, benefit, and motivation. *Neuron*, 112, 500–514.e5. <https://doi.org/10.1016/j.neuron.2023.10.038>
- Fearnley, J. M., & Lees, A. J. (1991). AGEING AND PARKINSON'S DISEASE: SUBSTANTIA NIGRA REGIONAL SELECTIVITY. *Brain*, 114, 2283–2301. <https://doi.org/10.1093/brain/114.5.2283>
- Fino, E., & Venance, L. (2011). Spike-timing dependent plasticity in striatal interneurons. *Neuropharmacology*, 60, 780–788. <https://doi.org/10.1016/j.neuropharm.2011.01.023>
- Frank, M. J. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Netw.*, 19, 1120–1136. <https://doi.org/10.1016/j.neunet.2006.03.006>
- Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *J Cogn Neurosci*, 17, 51–72. <https://doi.org/10.1162/0899929052880093>
- Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol Rev.*, 113, 300–326. <https://doi.org/10.1037/0033-295X.113.2.300>
- Frank, M. J., Seeberger, L. C., & O'reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science*, 306, 1940–1943. <https://doi.org/10.1126/science.1102941>
- Gerfen, C. R., & Surmeier, D. J. (2011). Modulation of striatal projection systems by dopamine. *Annu Rev Neurosci.*, 34, 441–466. <https://doi.org/10.1146/annurev-neuro-061010-113641>
- Grace, A. A. (2016). Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nat Rev Neurosci.*, 17, 524–532. <https://doi.org/10.1038/nrn.2016.57>
- Grace, A. A. (2001). Psychostimulant actions on dopamine and limbic system function: Relevance to the pathophysiology and treatment of ADHD. In *Stimulant Drugs and ADHD: Basic and Clinical Neuroscience* (pp. 134–157). New York, NY, US: Oxford University Press. <https://doi.org/10.1176/appi.ajp.159.11.1959-a>
- Grace, A. A. (1991). Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: A hypothesis for the etiology of schizophrenia. *Neuroscience*, 41, 1–24. [https://doi.org/10.1016/0306-4522\(91\)90196-u](https://doi.org/10.1016/0306-4522(91)90196-u)
- Graef, S., Biele, G., Krugel, L., Marzink, F., Wahl, M., Wotka, J., Klostermann, F., & Heekeren, H. (2010). Differential Influence of Levodopa on Reward-Based Learning in Parkinson's Disease. *Frontiers in Human Neuroscience*, 4.
- Grogan, J.P., Tsivos, D., Smith, L., Knight, B.E., Bogacz, R., Whone, A., Coulthard, E.J., 2017. Effects of dopamine on reinforcement learning and consolidation in parkinson's disease. *eLife* 6. DOI: 10.7554/eLife.26801.
- Gurney, K., Prescott, T. J., & Redgrave, P. (2001a). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biol Cybern.*, 84, 401–410. <https://doi.org/10.1007/PL00007984>
- Gurney, K., Prescott, T. J., & Redgrave, P. (2001b). A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour. *Biol Cybern.*, 84, 411–423. <https://doi.org/10.1007/PL00007985>
- Guthrie, M., Leblois, A., Garenne, A., & Borud, T. (2013). Interaction between cognitive and motor cortico-basal ganglia loops during decision making: A computational study. *J Neurophysiol.*, 109, 3025–3040. <https://doi.org/10.1152/jn.00026.2013>
- Guthrie, M., Myers, C. E., & Gluck, M. A. (2009). A neurocomputational model of tonic and phasic dopamine in action selection: A comparison with cognitive deficits in Parkinson's disease. *Behav Brain Res.*, 200, 48–59. <https://doi.org/10.1016/j.bbr.2008.12.036>
- Hollerman, J. R., Tremblay, L., & Schultz, W. (2000). Involvement of basal ganglia and orbitofrontal cortex in goal-directed behavior. *Prog Brain Res.*, 126, 193–215. [https://doi.org/10.1016/S0079-6123\(00\)26015-9](https://doi.org/10.1016/S0079-6123(00)26015-9)
- Humphries, M. D., Khamassi, M., & Gurney, K. (2012). Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. *Frontiers in Neuroscience*, 6.
- Humphries, M. D., Obeso, J. A., & Dreyer, J. K. (2018). Insights into Parkinson's disease from computational models of the basal ganglia. *J Neurol Neurosurg Psychiatry*, 89, 1181–1188. <https://doi.org/10.1136/jnnp-2017-315922>
- Humphries, M. D., Stewart, R. D., & Gurney, K. N. (2006). A physiologically plausible model of action selection and oscillatory activity in the basal ganglia. *J Neurosci.*, 26, 12921–12942. <https://doi.org/10.1523/JNEUROSCI.3486-06.2006>
- Kato, A., & Morita, K. (2016). Forgetting in Reinforcement Learning Links Sustained Dopamine Signals to Motivation. *PLoS Comput Biol.*, 12, e1005145.
- Kish, S. J., Shannak, K., & Hornykiewicz, O. (1988). Uneven Pattern of Dopamine Loss in the Striatum of Patients with Idiopathic Parkinson's Disease. *New England Journal of Medicine*, 318, 876–880. <https://doi.org/10.1056/NEJM198804073181402>
- Klein, M. O., Battagello, D. S., Cardoso, A. R., Hauser, D. N., Bittencourt, J. C., & Correa, R. G. (2019). Dopamine: Functions, Signaling, and Association with Neurological Diseases. *Cell Mol Neurobiol.*, 39, 31–59. <https://doi.org/10.1007/s10571-018-0632-3>
- Kruzich, P. J., Mitchell, S. H., Younkin, A., & Grandy, D. K. (2006). Dopamine D2 receptors mediate reversal learning in male C57BL/6J mice. *Cognitive, Affective, & Behavioral Neuroscience*, 6, 86–90. <https://doi.org/10.3758/CABN.6.1.86>
- Kwak, S., & Jung, M. W. (2019). Distinct roles of striatal direct and indirect pathways in value-based decision making. *eLife*, 8, e46050.
- Lawrence, A. D., Evans, A. H., & Lees, A. J. (2003). Compulsive use of dopamine replacement therapy in Parkinson's disease: Reward systems gone awry? *The Lancet Neurology*, 2, 595–604. [https://doi.org/10.1016/S1474-4422\(03\)00529-5](https://doi.org/10.1016/S1474-4422(03)00529-5)
- Liu, C., Wang, J., Li, H., Fietkiewicz, C., & Loparo, K. A. (2018). Modeling and Analysis of Beta Oscillations in the Basal Ganglia. *IEEE Trans Neural Netw Learn Syst.*, 29, 1864–1875. <https://doi.org/10.1109/TNNLS.2017.2688426>
- Maith, O., Baladron, J., Einhäuser, W., & Hamker, F. H. (2023). Exploration behavior after reversals is predicted by STN-GPe synaptic plasticity in a basal ganglia model. *iScience* 26, Article 106599. <https://doi.org/10.1016/j.isci.2023.106599>
- McCoy, B., Jahfari, S., Engels, G., Knapen, T., & Theeuwes, J. (2019). Dopaminergic medication reduces striatal sensitivity to negative outcomes in Parkinson's disease. *Brain*, 142, 3605–3620. <https://doi.org/10.1093/brain/awz276>
- Mikhail, J. G., Kim, H. R., Uchida, N., & Gershman, S. J. (2022). The role of state uncertainty in the dynamics of dopamine. *Current Biology*, 32, 1077–1087.e9. <https://doi.org/10.1016/j.cub.2022.01.025>
- Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Prog Neurobiol.*, 50, 381–425. [https://doi.org/10.1016/s0301-0082\(96\)00042-1](https://doi.org/10.1016/s0301-0082(96)00042-1)

- Morita, K., & Kato, A. (2022). Dopamine ramps for accurate value learning under uncertainty. *Trends in Neurosciences*, *45*, 254–256. <https://doi.org/10.1016/j.tins.2022.01.008>
- Morita, K., Kato, A., 2018. A Neural Circuit Mechanism for the Involvements of Dopamine in Effort-Related Choices: Decay of Learned Values, Secondary Effects of Depletion, and Calculation of Temporal Difference Error. *eNeuro* 5, ENEURO.0021-18.2018. DOI: 10.1523/ENEURO.0021-18.2018.
- Moustafa, A. A., Cohen, M. X., Sherman, S. J., & Frank, M. J. (2008). A role for dopamine in temporal decision making and reward maximization in parkinsonism. *J. Neurosci.*, *28*, 12294–12304. <https://doi.org/10.1523/JNEUROSCI.3116-08.2008>
- Moustafa, A. A., & Gluck, M. A. (2011). A neurocomputational model of dopamine and prefrontal-striatal interactions during multicue category learning by Parkinson patients. *J Cogn Neurosci*, *23*, 151–167. <https://doi.org/10.1162/jocn.2010.21420>
- Moustafa, A.A., Herzallah, M.M., Gluck, M.A., 2014. A model of reversal learning and working memory in medicated and unmedicated patients with Parkinson's disease. *Journal of Mathematical Psychology, Special Issue in Honor of William K. Estes* 59, 120–131. DOI: 10.1016/j.jmp.2013.08.003.
- Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology (Berl)*, *191*, 507–520. <https://doi.org/10.1007/s00213-006-0502-4>
- O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput*, *18*, 283–328. <https://doi.org/10.1162/089976606775093909>
- Parent, M., Wallman, M.-J., Gagnon, D., & Parent, A. (2011). Serotonin innervation of basal ganglia in monkeys and humans. *J Chem Neuroanat*, *41*, 256–265. <https://doi.org/10.1016/j.jchemneu.2011.04.005>
- Rigoli, F., Chew, B., Dayan, P., & Dolan, R. J. (2016). Multiple value signals in dopaminergic midbrain and their role in avoidance contexts. *Neuroimage*, *135*, 197–203. <https://doi.org/10.1016/j.neuroimage.2016.04.062>
- Rigoli, F., Rutledge, R. B., Chew, B., Ousdal, O. T., Dayan, P., & Dolan, R. J. (2016). Dopamine Increases a Value-Independent Gambling Propensity. *Neuropsychopharmacology*, *41*, 2658–2667. <https://doi.org/10.1038/npp.2016.68>
- Rinne, J. O. (1993). Nigral degeneration in Parkinson's disease. *Mov Disord*, *8*(Suppl 1), S31–S35. <https://doi.org/10.1002/mds.870080507>
- Robinson, O. J., Frank, M. J., Sahakian, B. J., & Cools, R. (2010). Dissociable responses to punishment in distinct striatal regions during reversal learning. *Neuroimage*, *51*, 1459–1467. <https://doi.org/10.1016/j.neuroimage.2010.03.036>
- Rutledge, R. B., Lazzaro, S. C., Lau, B., Myers, C. E., Gluck, M. A., & Glimcher, P. W. (2009). Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *Journal of Neuroscience*, *29*, 15104–15114. <https://doi.org/10.1523/JNEUROSCI.3524-09.2009>
- Sala-Bayo, J., Fiddian, L., Nilsson, S. R. O., Hergiv, M. E., McKenzie, C., Mareschi, A., Boulos, M., Zhukovsky, P., Nicholson, J., Dalley, J. W., Alsió, J., & Robbins, T. W. (2020). Dorsal and ventral striatal dopamine D1 and D2 receptors differentially modulate distinct phases of serial visual reversal learning. *Neuropsychopharmacology*, *45*, 736–744. <https://doi.org/10.1038/s41386-020-0612-4>
- Salinas, A. G., Lee, J. O., Augustin, S. M., Zhang, S., Patriarchi, T., Tian, L., Morales, M., Mateo, Y., & Lovinger, D. M. (2023). Distinct sub-second dopamine signaling in dorsolateral striatum measured by a genetically-encoded fluorescent sensor. *Nat Commun*, *14*, 5915. <https://doi.org/10.1038/s41467-023-41581-3>
- Saunders, B. T., Richard, J. M., Margolis, E. B., & Janak, P. H. (2018). Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. *Nat Neurosci*, *21*, 1072–1083. <https://doi.org/10.1038/s41593-018-0191-4>
- Schirru, M., Véronneau-Veilleux, F., Nekka, F., & Ursino, M. (2022). Phasic Dopamine Changes and Hebbian Mechanisms during Probabilistic Reversal Learning in Striatal Circuits: A Computational Study. *Int J Mol Sci*, *23*, 3452. <https://doi.org/10.3390/ijms23073452>
- Schroll, H., Beste, C., & Hamker, F. H. (2015). Combined lesions of direct and indirect basal ganglia pathways but not changes in dopamine levels explain learning deficits in patients with Huntington's disease. *Eur J Neurosci*, *41*, 1227–1244. <https://doi.org/10.1111/ejn.12868>
- Schroll, H., & Hamker, F. H. (2013). Computational models of basal-ganglia pathway functions: Focus on functional neuroanatomy. *Front Syst Neurosci*, *7*, 122. <https://doi.org/10.3389/fnsys.2013.00122>
- Schultz, W. (2016). Dopamine reward prediction error coding. *Dialogues Clin Neurosci*, *18*, 23–32.
- Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J Neurophysiol*, *80*, 1–27. <https://doi.org/10.1152/jn.1998.80.1.1>
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599. <https://doi.org/10.1126/science.275.5306.1593>
- Suryanarayana, S. M., Hellgren Kotaleski, J., Grillner, S., & Gurney, K. N. (2019). Roles for globus pallidus externa revealed in a computational model of action selection in the basal ganglia. *Neural Netw*, *109*, 113–136. <https://doi.org/10.1016/j.neunet.2018.10.003>
- 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*, 596–612. [https://doi.org/10.1016/S0028-3932\(99\)00103-7](https://doi.org/10.1016/S0028-3932(99)00103-7)
- Ursino, M., & Baston, C. (2018). Aberrant learning in Parkinson's disease: A neurocomputational study on bradykinesia. *Eur. J. Neurosci.*, *47*, 1563–1582. <https://doi.org/10.1111/ejn.13960>
- Ursino, M., Magosso, E., Lopane, G., Calandra-Buonaura, G., Cortelli, P., & Contin, M. (2020). Mathematical modeling and parameter estimation of levodopa motor response in patients with parkinson disease. *PLoS ONE*, *15*, e0229729.
- Ursino, M., Véronneau-Veilleux, F., & Nekka, F. (2020). A non-linear deterministic model of action selection in the basal ganglia to simulate motor fluctuations in Parkinson's disease. *Chaos*, *30*, Article 083139. <https://doi.org/10.1063/5.0013666>
- van Albada, S. J., & Robinson, P. A. (2009). Mean-field modeling of the basal ganglia-thalamocortical system. I: Firing rates in healthy and parkinsonian states. *Journal of Theoretical Biology*, *257*, 642–663. <https://doi.org/10.1016/j.jtbi.2008.12.018>
- van Swieten, M. M. H., & Bogacz, R. (2020). Modeling the effects of motivation on choice and learning in the basal ganglia. *PLoS Comput Biol*, *16*, e1007465.
- Verharen, J. P. H., Adan, R. A. H., & Vanderschuren, L. J. M. J. (2019). Differential contributions of striatal dopamine D1 and D2 receptors to component processes of value-based decision making. *Neuropsychopharmacology*, *44*, 2195–2204. <https://doi.org/10.1038/s41386-019-0454-0>
- Véronneau-Veilleux, F., Robaey, P., Ursino, M., & Nekka, F. (2022). A mechanistic model of ADHD as resulting from dopamine phasic/tonic imbalance during reinforcement learning. *Front Comput Neurosci*, *16*, Article 849323. <https://doi.org/10.3389/fncom.2022.849323>
- Véronneau-Veilleux, F., Robaey, P., Ursino, M., & Nekka, F. (2021). An integrative model of Parkinson's disease treatment including levodopa pharmacokinetics, dopamine kinetics, basal ganglia neurotransmission and motor action throughout disease progression. *J Pharmacokinetic Pharmacodyn*, *48*, 133–148. <https://doi.org/10.1007/s10928-020-09723-y>
- Voon, V., Fernagut, P.-O., Wickens, J., Baunez, C., Rodriguez, M., Pavon, N., Juncos, J. L., Obeso, J. A., & Bezaud, E. (2009). Chronic dopaminergic stimulation in Parkinson's disease: From dyskinesias to impulse control disorders. *The Lancet Neurology*, *8*, 1140–1149. [https://doi.org/10.1016/S1474-4422\(09\)70287-X](https://doi.org/10.1016/S1474-4422(09)70287-X)
- Wang, Y., Toyoshima, O., Kunimatsu, J., Yamada, H., & Matsumoto, M. (2021). Tonic firing mode of midbrain dopamine neurons continuously tracks reward values changing moment-by-moment. *Elife*, *10*, e63166.
- Westbrook, A., Frank, M. J., & Cools, R. (2021). A mosaic of cost-benefit control over cortico-striatal circuitry. *Trends Cogn Sci*, *25*, 710–721. <https://doi.org/10.1016/j.tics.2021.04.007>
- Zénon, A., Devesse, S., & Olivier, E. (2016). Dopamine Manipulation Affects Response Vigor Independently of Opportunity Cost. *J Neurosci*, *36*, 9516–9525. <https://doi.org/10.1523/JNEUROSCI.4467-15.2016>