Action Learning in an Integrated Model of Basal Ganglia and its Application in Control Systems

Álvaro González Redondo

Department of Computers Engineering, Automatics and Robotics (ICAR)

Doctoral Program in Information and Communication Technologies

University of Granada

A dissertation submitted for the degree of *Philosophiae Doctor (PhD) in Information and Communication Technologies*

Nov 2023

Supervisors: Eduardo Ros Vidal and Jesús A. Garrido Alcázar

UNIVERSIDAD DE GRANADA

Action Learning in an Integrated Model of Basal Ganglia and its Application in Control Systems

(Aprendizaje de Acciones en un Modelo Integrado de Ganglios Basales y su Aplicación en Sistemas de Control)

Dissertation presented by:
Álvaro González-Redondo

To apply for the:
International PhD degree in Information and Communication Technologies

Signed: Álvaro González Redondo

Editor: Universidad de Granada. Tesis Doctorales
Autor: Álvaro González Redondo
ISBN: 978-84-1195-191-3
URI: https://hdl.handle.net/10481/89805

Agradecimientos

Quiero comenzar reconociendo a mis colegas de trabajo, quienes han sido pilares en esta etapa. A Eduardo, por creer en mí cuando yo mismo dudaba; a Jesús, por su paciencia y por compartir su invaluable experiencia; a Paco, siempre dispuesto a ofrecer su ayuda; y a Niceto, por procurar un ambiente familiar en el entorno laboral, y por su honestidad tan implacable como reveladora. A Javier, por transmitirme su entusiasmo; a Luis, por las conversaciones y risas que aligeraron muchos de mis días; y a Mila y a todos aquellos que han contribuido a crear buenos momentos.

No puedo dejar de mencionar a Zahra y a los compañeros de la oficina en Estocolmo, cuyas invitaciones a chocolate caliente se convirtieron en uno de los momentos más esperados de aquellos días. A Jeanette y Sten, por su escucha activa y sus constructivas sugerencias que durante años han enriquecido mi trabajo.

A Jaime, por ser lo que se espera de un hermano, y serlo independientemente de las circunstancias.

A Lara, por ser; a Blanca, Santi y Nani, por estar; y a Lenka, por volver.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ACh: Acetylcholine

BG: Basal ganglia

DA: Dopamine

GPe: Globus pallidus externa

LIF: Leaky integrate-and-fire

MSNs: Medium spiny neurons

RL: Reinforcement learning

SNc: Substantia nigra pars compacta

SNNs: Spiking neural networks

SNr: Substantia nigra pars reticulata

STDE: Spike-timing-dependent eligibility

STDP: Spike-timing-dependent plasticity

STN: Subthalamic nucleus

STR: Striatum

Abstract

This thesis is motivated by the need to comprehend the complex functionalities of the nervous system, specifically the basal ganglia (BG), due to its significant role in neurological disorders and behavioral processes. Traditional experimental approaches have fallen short in explaining the contributions of brain structures to intricate behaviors, thus necessitating alternative approaches like computational modeling. This thesis aims to bridge this gap by employing biologically-inspired computational models, with a focus on spiking neural networks (SNNs) due to their ability to emulate the temporal dynamics of biological neurons, to simulate the basal ganglia. The simulation aims to understand the basal ganglia's role in action selection and the influence of neuromodulators, particularly dopamine and acetylcholine, on learning and decision-making. The ultimate objective is to link and integrate insights from neuroscience to embodied agents, by applying the knowledge of motor learning processes within the basal ganglia to the development of action selection in control systems.

This thesis presents several contributions to the field of computational neuroscience, focusing on models of the basal ganglia. First, it introduces a computational model to elucidate the paradoxical sensorial improvement observed in patients with Huntington's disease, suggesting that both dopamine levels and the early stage of affliction may independently play significant roles. Secondly, it undertakes a thorough examination of methods for tuning spiking neural models of striatum plasticity, highlighting the effectiveness of automatic optimization algorithms in calibrating the models. Third, it constructs a biologically-inspired network model of the striatum integrating features such as spike-timing-dependent plasticity, homeostatic mechanisms, and lateral inhibitory connections, capable of recognizing complex patterns and choosing rewarded actions. Additionally, it develops a functional striosome model for reward prediction error (RPE) in the basal ganglia. Finally, it refines an existing striatal reinforcement learning model by incorporating acetylcholine as a local population feedback, which demonstrates proficiency in pattern recognition and action selection, while offering insights into the brain's learning mechanisms and the role of neuromodulators.

The research presented in this thesis has explored the neural mechanisms involved in action selection, learning, and decision-making, with a particular focus on the striatum and the basal ganglia. Through computational models and novel methodologies, this work has contributed to an enhanced understanding of how neuronal populations and neuromodulators interact within the basal ganglia, and potentially with other brain regions. While the insights gained are promising, the study acknowledges certain limitations, notably the sensitivity of the model's internal dynamics to the form of input representation. Future research is encouraged to further validate these findings and explore additional biological factors, such as the basal ganglia-cortex loop and the role of interneurons. Moreover, the use of high-performance computing platforms and the development of novel optimization techniques are suggested as avenues for refining the computational models. This could have practical applications, including advancing knowledge on neurological disorders like Huntington's disease and fostering the development of bio-inspired reinforcement agents.

Resumen

Esta tesis está motivada por la necesidad de comprender las complejas funcionalidades del sistema nervioso, específicamente los ganglios basales (GB), debido a su importante papel en trastornos neurológicos y procesos de comportamiento. Los enfoques experimentales tradicionales no han logrado explicar satisfactoriamente cómo las estructuras cerebrales contribuyen a comportamientos complejos, lo cual hace necesaria la adopción de enfoques alternativos como el modelado computacional. Esta tesis tiene como objetivo llenar este vacío mediante el uso de modelos computacionales inspirados biológicamente, centrados principalmente en redes neuronales de impulsos (SNNs), para simular los ganglios basales. La simulación se enfoca en entender el papel de los ganglios basales en la selección de acciones y la influencia de neuromoduladores, particularmente la dopamina y la acetilcolina, en el aprendizaje y la toma de decisiones. El objetivo final es vincular e integrar perspectivas de la neurociencia con agentes corpóreos, aplicando el conocimiento de los procesos de aprendizaje motor dentro de los ganglios basales al desarrollo de selección de acciones en sistemas de control.

Esta tesis presenta varias contribuciones al campo de la neurociencia computacional, centrándose en modelos de los ganglios basales. En primer lugar, introduce un modelo computacional para esclarecer la paradójica mejora sensorial observada en pacientes con enfermedad de Huntington, sugiriendo que los niveles de dopamina y las etapas tempranas de la afección pueden desempeñar un papel significativo. En segundo lugar, realiza un examen exhaustivo de la optimización de modelos neuronales de spikes del estriado para la plasticidad, destacando la efectividad de los algoritmos de optimización automática en el ajuste de modelos. En tercer lugar, construye un modelo de red biológicamente inspirado del estriado, integrando características como la plasticidad dependiente del tiempo de disparo, mecanismos homeostáticos y conexiones inhibitorias laterales, capaz de reconocer patrones complejos y elegir acciones recompensadas. Además, desarrolla un modelo funcional de estriosoma para el error de predicción de recompensa (RPE) en los ganglios basales. Finalmente, refina un modelo existente de aprendizaje por refuerzo en el estriado incorporando la acetilcolina como retroalimentación de la población local, lo que demuestra habilidad en el reconocimiento de patrones y selección de acciones, a la vez que ofrece ideas sobre los mecanismos de aprendizaje del cerebro y el papel de los neuromoduladores.

La investigación presentada en esta tesis ha explorado los mecanismos neuronales involucrados en la selección de acciones, el aprendizaje y la toma de decisiones, con un enfoque particular en el estriado y los ganglios basales. A través de modelos computacionales y metodologías novedosas, este trabajo ha contribuido a una mejor comprensión de cómo las poblaciones neuronales y los neuromoduladores interactúan dentro de los ganglios basales. Si bien los conocimientos adquiridos son prometedores, el estudio reconoce ciertas limitaciones, en particular, la sensibilidad de la dinámica interna del modelo a la forma de representación de la entrada. Se alienta a investigaciones futuras a validar y ampliar estos hallazgos, y explorar factores biológicos adicionales, como el bucle ganglios basales-corteza y el papel de las interneuronas. Además, se sugiere el empleo de

plataformas de computación de alto rendimiento y el desarrollo de nuevas técnicas de optimización como vías para refinar aún más estos modelos. Esto podría tener aplicaciones prácticas, como avanzar en el conocimiento sobre trastornos neurológicos como la enfermedad de Huntington y fomentar el desarrollo de agentes de refuerzo bioinspirados.

Chapter 1: Introduction

1. The Importance of Understanding the Brain and Its Applications

Understanding the workings of the nervous system and its related pathologies is one of the most important challenges of this century. It is estimated that nearly 300 million people suffer from various types of neurological disorders each year in Europe, with an annual economic cost of 800,000 million euros, and 83,000 million just in Spain (Parés-Badell et al. 2014).

Traditional research on the brain has used both in vivo and in vitro experiments to study natural and artificially induced alterations in brain circuitry. In-vivo experiments involve living organisms, while in-vitro experiments take place in controlled artificial environments where simpler relationships can be studied with more clarity. These experiments have provided valuable insights into how specific cells interact and how diseases and treatments affect the brain's normal operation. However, due to the complex nature of brain function, understanding its role in various behaviors remains highly elusive.

Computational modeling of neural systems, based on mathematical models and experimental data obtained from computer simulations (in-silico), provides a promising approach to address the limitations of traditional experimental techniques in understanding complex brain behaviors. Selecting the appropriate level of abstraction for simulating brain components is crucial in these models. Neurons, the fundamental processing units in the nervous system, can be modeled using simplified integrate-and-fire models, which account for the accumulation of input activity in their membrane potential. In addition, synapses play a vital role in these models as they facilitate communication between neurons, regulating the transmission of signals from one cell to another.

The approach employed in this thesis involves simulating networks comprising modeled neurons and synapses as the basic processing units. By doing so, it is possible to construct more complex systems (nervous centers) that **allow us to experiment and validate working hypotheses** by comparing the simulated results against findings in real biological systems. The use of biologically inspired computational models that simulate SNNs has been shown to be effective in understanding experimental recordings from multiple brain areas (Ghosh-Dastidar and Adeli 2007; 2009b) and in studying different neurological alterations (Geminiani et al. 2018; Antunes, Faria da Silva, and Simoes de Souza 2018; Antonietti et al. 2018).

Studying the brain not only facilitates understanding biological systems, but it also has **practical applications, such as the improvement of human inspired robotics**. While significant progress has been made in artificial intelligence for image (Krizhevsky, Sutskever, and Hinton 2017), speech recognition (Dahl et al. 2012), and translation and text generation (Vaswani et al. 2017), robotics has not yet achieved the same level of flexibility in learning new tasks as biological systems due to limitations in algorithms and techniques. This limitation has confined robotics primarily to the industrial sector (highly repetitive tasks in structured environments). By understanding how the brain, specifically structures like the basal ganglia, solve problems and learn action selection, we can

potentially transfer these principles to develop more efficient and adaptable robotic control systems (Krichmar 2018). This is particularly relevant as reinforcement learning algorithms, which offer the most applied approach in this framework, but these algorithms still require large amounts of difficult and costly data, among other challenges (Dulac-Arnold, Mankowitz, and Hester 2019).

2. Motivation

In computational neuroscience, we build models to evaluate and answer specific questions related to the brain's functioning. As stated by the Human Brain Project Work Package 3, in which this work has been partially framed: *"The central ambition of the Work Package consists in achieving a measurable step forward in our understanding of human cognition; specifically, how biological learning networks enable human visuo-motor and cognitive functions"*. A key challenge is understanding how the basal ganglia can solve reinforcement learning problems via networks with segregated action requests, complex dynamics, and multiple neuromodulator interactions. It is also essential to explore how Huntington's disease affects the information processing capabilities of this structure. Therefore, we aim to create basal ganglia models that offer insights into these aspects.

Additionally, this Work Package encourages specific approaches, primarily focusing on numerical simulations, embodiment, and real-world systems. In this thesis, we develop and simulate network models that incorporate functionally relevant features and simulate abstract tasks that resemble real tasks. Future work includes studying these models within physically realistic tasks and examining the impact of different related brain areas (such as the cerebellum) on motor learning and overall basal ganglia functioning. By employing network models capable of solving reinforcement learning tasks, we explore the functional implications of the neuronal system. This thesis serves as a first step towards better understanding the brain's fundamentals and mechanisms of neurological diseases.

3. Objectives

This thesis is focused on developing biologically plausible computational models of the basal ganglia, with the intention of furthering our understanding of motor learning processes (specifically action selection) and enabling their application to embodied systems. In order to accomplish this primary objective, several specific objectives were identified and addressed through the work detailed in this thesis:

1. Investigating basal ganglia-related brain pathologies, such as Huntington's disease, using computational models. We aim to study the effects of changes in dopamine levels and the influence of Huntington's disease (in which dopamine levels increase with respect to the general standard population), the goal is to better understand information processing and basal ganglia dynamics.
2. Simulating Basal Ganglia models in action selection and reinforcement learning tasks. The thesis seeks to create a computational model of the striatum in Basal Ganglia that incorporates multiple biologically-inspired mechanisms in order to understand their

interactions during normal functioning within a reinforcement learning task.

3. Examining the interaction between the Basal Ganglia model and other brain areas, as well as their impact on motor learning in decision-making tasks. This aspect of the research aims to improve our understanding of how various brain structures interact and influence decision-making processes.

In summary, these objectives aim to contribute to the advancement of knowledge regarding basal ganglia function, the role of neuromodulators, and the relationships between these factors and neurological disorders. It is expected that these insights will ultimately help elucidate the neural basis of decision-making and learning processes. This, in turn, may lead to two potential outcomes: on the one hand, the development of novel approaches for understanding brain pathologies related to the basal ganglia, and on the other hand, the creation of bioinspired control systems for embodied agents.

4. Thesis Overview and Chapter Organization

This thesis primarily aims to explore the brain's neural mechanisms and processes that are involved in making decisions, choosing actions, learning, and also how they are affected in disorders like Huntington's disease. This is done by creating computational models that focus on the basal ganglia (BG) and its internal structures, including the striatum. The models combine several biology-based mechanisms and important neuromodulators to give a clearer picture of the complex functions of the brain.

The thesis is structured into the following chapters:

Chapter 1: Explains the motivation, objectives, and general contributions of the thesis.

Chapter 2: Introduces the field, explaining basic concepts and reviewing existing research.

Chapter 3: Presents the results of the works carried out during the thesis, including both published (added in the annex) and unpublished works, and delves into more detailed aspects of the unpublished works.

- First, we presented a work that looks into the unexpected improvement in some tasks observed in patients with Huntington's disease (González-Redondo et al. 2020). The computational model developed offers insights into how changes in dopamine levels and the effects of Huntington's disease can impact the processing of information in the BG. This helps in understanding why these patients show better performance in certain discrimination tasks and provides information on the fast-acting functions of the basal ganglia.
- Next, optimization algorithms in automatically fine-tuning spiking neural models (Cruz et al. 2022) are also considered.
- In the last included publication, the role of the striatum in reinforcement learning and action selection (González-Redondo et al. 2023) is discussed. The computational model incorporates spike-timing-dependent plasticity (STDP), homeostatic mechanisms, and

asymmetric lateral inhibitory connections. This model successfully learns and selects actions that are most rewarding. This research also highlights how different neural and network features contribute to making effective decisions, emphasizing the significance of the striatum structure and its internal dynamics in this function.

- The unpublished work titled “Striosome Model for Reward Prediction Error” (section 2.1.) introduces a model focusing on reward prediction error (RPE) within the basal ganglia’s striosomes, offering insights into how these structures are involved in processing rewards and motivating behavior.
- The work titled “Dopamine and Acetylcholine Modulation in a Reinforcement Learning Striatal Model” (section 2.2.), attention is given to the role of neuromodulators dopamine (DA) and acetylcholine (ACh) function in the striatum. The study uses a reinforcement learning computational model to show the relationship between stimulus and action. Including DA and ACh, along with lateral connections and homeostatic mechanisms, the model illustrates how these neuromodulators and network components facilitate efficient pattern recognition and adaptation to changes. Furthermore, it highlights the specific contribution of incorporating ACh feedback in speeding up the learning process and facilitating scalability to more complex decisions.

Chapter 4: Summarizes the significance of the results obtained, enumerates the main contributions in detail, and discusses future work.

Annex: Includes journal articles that compose the primary outcomes of this thesis.

To sum up, this thesis is an in-depth study of the basal ganglia's role, structures and processes. It focuses on integrating biology-based mechanisms, reinforcement learning, and neuromodulators to better understand how actions are selected and how learning takes place. Furthermore, it helps understanding how the performance of certain tasks can be improved, especially when disorders like Huntington's disease are present. The research in this thesis contributes to the ongoing efforts to understand specific aspects of the brain's functions, particularly within the basal ganglia. By focusing on particular neural mechanisms and their roles in behavior, learning, and the effects of certain disorders, this thesis adds to the growing body of knowledge in this specialized area.

5. Research Projects and Funding Framework

The work presented has been conducted within the scope of the national project INTSENSO (Integración Sensorimotora para control adaptativo mediante aprendizaje en cerebelo y centros nerviosos relacionados. Aplicación en robótica.) (MICINN-FEDER-PID2019-109991GB-I00), as well as the European Human Brain Project (HBP) (SGA3) (H2020 SGA3. 945539), primarily focusing on the development of functional models in Work Package 3. The research was specifically supported by the FPU grant (FPU17/04432).

The international framework of the HBP facilitated collaborations with other European research teams

that have extensive expertise in this field, as evidenced by the co-authorship of some publications. Additionally, the HBP enabled a 3-month research stay at KTH Royal Institute of Technology with the research group led by Jeanette Hellgren Kotaleski and Sten Grillner in Stockholm, Sweden. This enables the "International Mention" of the PhD.

Chapter 2: Thesis contextualization

While numerous questions about the functioning of the basal ganglia remain unanswered, years of research have yielded certain structural and functional principles. Our models build upon these principles, and as such, it will be beneficial for the reader to be familiar with these general concepts. In this chapter, we provide a concise overview of the relevant principles and their relation to previous research.

1. Basal Ganglia

1.1. Roles of the Basal Ganglia

Choosing the right action among many available choices represents a primary but also challenging behavior for animal species. Multiple stimuli spanning different sensory modalities continuously converge to the brain, and adequate responses (taking into account these inputs) need to be decided, often as fast as possible. For example, a rat could spot a hidden piece of food, and at the same time smell the scent of a nearby predator, feel the tiredness in its limbs, and stomach pangs of hunger. This decision process usually requires the animals to ignore most of the information available and focus on what is relevant now. To be successful, many actions must be performed serially, which requires prioritization and temporal organization of the behavior; the rat cannot simultaneously run towards the food, catch it, find a safe place, and eat it – they have to be done one after the other. Finally, the consequences (good or bad) of all these decisions need to be remembered, in order to make better decisions in the future and to be able to avoid fatal mistakes.

The **basal ganglia (BG)** is the part of the brain more implicated in these roles. They are a collection of highly interconnected nuclei located in the deepest part of the brain. For a long time the role of the BG was not clear as they seemed to cover many different behaviors. It has long been known its implication in diseases such as Parkinson's or Huntington's or behavioral disorders like Tourette syndrome or obsessive compulsive disorder (Obeso, Olanow, and Nutt 2000; Chesselet and Delfs 1996; Wolf et al. 1996; Maia, Cooney, and Peterson 2008). Clinical and theoretical studies have pointed to a variety of functions like the initiation of voluntary movements, the learning of habitual movements or to temporally organize the behavior (Jankovic 2008; Peter Redgrave et al. 2010; Harrington, Haaland, and Hermanowitz 1998). More general roles that could explain all the previous ones have been proposed, like “chunking” or grouping common motor sequences into unconsciously manageable entities (Graybiel 1998), or to minimize prediction errors, both in learning and planning (Bogacz 2020). Biological studies (Grillner et al. 2005; Graybiel 1998; Hikosaka, Takikawa, and Kawagoe 2000) and relevant computational models (K. Gurney, Prescott, and Redgrave 2001) also have proposed the association between the BG and the action-selection and reinforcement learning. Specifically, the cortex and other brain structures send action proposals to the basal ganglia, which select the appropriate ones to be executed in the current context. Past experiences weigh the selection favoring the ones that resulted better in similar contexts.

1.2. Functioning of the Basal Ganglia

Structure is related to function in the brain [although this relationship is not simple (Suárez et al. 2020)], so it is useful to study its anatomy to better understand how it works. The BG network presents complex **anatomical and functional subdivisions**, but it is usually structured in five main neuron populations (Shipp 2017) which can be organized into the following three sections:

- The inputs of the BG are mainly received through the corpus striatum (STR), with its main cell type being the medium spiny neurons (MSNs), and the subthalamic nucleus (STN) neurons.
- The intermediate layers are composed by the external segment of the globus pallidus (GPe) and the substantia nigra pars compacta (SNc).
- The output projection to the thalamus is finally carried by the substantia nigra pars reticulata (SNr).

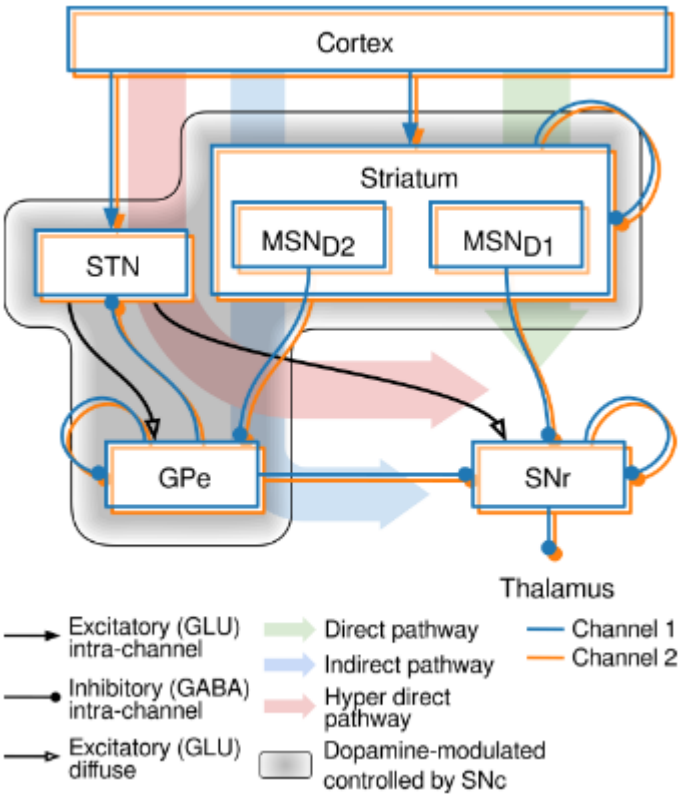

Figure 1**:** Basal ganglia structures and connectivity with different channels (blue and orange, see Channel Structure section) showing the direct, indirect and hyper direct pathways.

The **connectivity** of these populations is mainly drawn according to three main routes from the cortex to the thalamus as follows (Fig. 1):

- The **direct pathway**, where the cerebral cortex makes excitatory glutamatergic synapses into the MSN D1, which inhibits the SNr.

- The **indirect pathway**, where the cerebral cortex excites the MSN D2, which inhibits the GPe, and finally, the GPe which also inhibits the SNr.
- The **hyper direct pathway**, where the cortex makes glutamatergic connections into the STN, which diffusely excites the SNr.

The direct and indirect pathways are traditionally considered to **promote and inhibit behavior**, respectively. This works by selective disinhibition or inhibition, respectively, of targets in motor thalamus: as SNr is tonically active, its target (the thalamus) is normally inhibited. The activation of the direct pathway inhibits SNr, which results in disinhibition of the thalamus, allowing it to carry out actions. The indirect pathway instead inhibits the GPe which is also tonically active inhibiting the SNr, so it has the net effect of reinforcing the inhibition of the thalamus, reducing the behavior. Recent genetic and optical studies on striatal circuits have facilitated the testing of classical ideas about the functioning of this system. However, new models are necessary for a better understanding of the striatum's role in learning and decision-making (Cox and Witten 2019). Our research in this regard is presented in González-Redondo et al. (2020), which is included as an annex in this thesis.

In addition to these broad pathways, there are important neuromodulators like **Dopamine (DA) and acetylcholine (ACh)** influencing different parts of the circuitry. For example, there are dopaminergic projections from the SNc to the MSN, the STN and the GPe with modulatory effects (shaded box in Fig. 1). Also, thalamic projections innervate cholinergic interneurons in the STR, influencing the amount of ACh in this nucleus (Xiao and Roberts 2021). Phasic DA is generally thought to carry reinforcement-related signals to the STR (Hart et al. 2014) (although its full role is still debated (Berke 2018)), while ACh pauses seem to define the time window for phasic dopamine to induce plasticity (Reynolds et al. 2022). However, it remains under discussion how these mechanisms combine to make the striatum able to solve action-selection problems. Our investigation on this topic is presented in Chapter 3, Section *2.2. Dopamine and Acetylcholine Modulation in an Reinforcement Learning Striatal Model* of this thesis.

1.3. Learning in the Basal Ganglia

Learning is an important feature in the BG functioning. Animals are born with a set of genetically defined behaviors but most of them require refinement through learning (Brainard and Doupe 2002). Additionally, higher mammals and certain birds, such as primates and crows, are known to exhibit the capacity to acquire new behaviors through experience. Nevertheless, the specific role of the basal ganglia in such complex learning is still an area of research.

A fundamental form of learning involves responding to rewards and punishments: if an animal in a given situation takes an action that is shortly followed by a reward, that action is likely to be taken more often in similar situations. This ability to adapt behavior based on the rewards and punishments received is known as reinforcement learning (Sutton and Barto 2018). The BG is well situated for reinforcement learning as it serves as a neural interface between reinforcement signals, primarily through DA, and action representations via cortical input pathways (Mogenson, Jones, and Yim

1980). The STR and its main population of MSNs constitute this interface, where cortical inputs establish plastic synapses modulated by DA. The adjustment of the weights of these cortico-striatal synapses in response to reward signals influences which actions are prioritized in the future (Reynolds and Wickens 2002; K. N. Gurney, Humphries, and Redgrave 2015). Building upon these concepts, we conducted research that is presented in González-Redondo et al. (2023), and included as an annex.

2. Computational Models of the Basal Ganglia

2.1. Channel Structure

Several computational models of the BG have been created in the past years. One of the most influential models (K. Gurney, Prescott, and Redgrave 2001) tries to explain why the information flows segregated through the BG circuits (DeLong, Crutcher, and Georgopoulos 1985; Parent and Hazrati 1995). This is an important working principle of the BG, as it suggests how it processes the inputs it receives, and what its output means. It has been proposed that the BG processes a large number of cognitive streams or **channels** in parallel (K. Gurney, Prescott, and Redgrave 2001), each of them representing a feasible action to be performed (Suryanarayana et al. 2019). The BG are thought to act as an action selection machinery by inhibiting every nonselected action in the thalamus with the SNr, based on their corresponding activity level or salience (P. Redgrave, Prescott, and Gurney 1999). According to recent research, this segregation through the entire cortical-BG-thalamic loop shows a very high specificity to almost neuron-to-neuron level (Hunnicutt et al. 2016; Foster et al. 2021), which could mean that it seems feasible to impact behavior at different levels of detail.

2.2. Action Selection

Earlier models of BG tend to be population models, where each node represents a population of neurons instead of single neurons. The activity of these nodes does not describe individual spikes but rather the average activity of populations of neurons. There are more detailed models known as Spiking Neural Networks (SNNs) that model individual neurons that use spikes to compute and transmit information (Ghosh-Dastidar and Adeli 2009a). As the specific timing of spikes carry relevant information in many biological contexts (Maass 1997), these models are useful to understand how the brain computes at the neuronal description level. This is relevant for example to better understand how these channels interact between them during the action-selection process. Burke et al. (Burke, Rotstein, and Alvarez 2017) proposed a model of **asymmetric lateral connectivity in the STR** that tries to explain how different clusters of striatal neurons interact and which role they play in information processing [Fig. 5E in (Burke, Rotstein, and Alvarez 2017), and adapted here in Fig. 2]. This model offers an explanation for the in-vivo phenomenon of coactivation of subpopulations of D1 or D2 MSNs, which seems paradoxical as each subpopulation projects to behaviorally opposite pathways (direct and indirect, respectively). This structured connectivity pattern is determined by lateral inhibition between neurons that belong to the same channel and between neurons within

different channels but accounting for the same receptor type (D1 or D2). The authors also include asymmetrical connections with more intensive intra-channel inhibition from D2 to D1 neurons than in the opposite direction. This pattern resulted in synchronized phase-dependent activation between MSN D1 and D2 neuron groups that belong to different channels. This behavior has to be taken into account if we want to design models capable of action-selection.

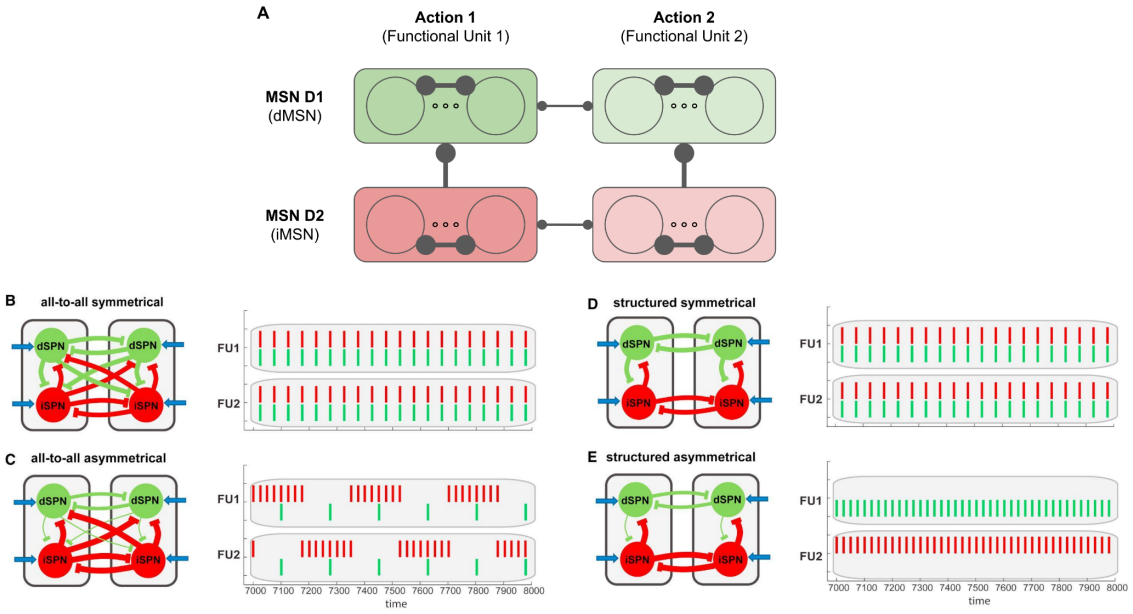

Figure 2**: A.** Connectivity pattern used in this thesis for modeling the channels in STR. Each column represents an action channel composed of D1 and D2 MSN subpopulations (in parenthesis the equivalent name given in Burke et al. (2017) work). There is inhibition within every subpopulation, and from D2 to D1 subpopulation within the same channel. Between different channels, only subpopulations of the same type have lateral inhibition. Inhibitory synapses are weak in all cases except in the case of D2 to D1 inhibition within the same channel. **B-E.** Figures adapted from Burke et al. (2017) where the resulting activity of different connectivity patterns are shown. The connectivity pattern used in this thesis corresponds to the structural asymmetrical pattern (E), where the lateral inhibition from the active channels limits activity of the other silent channels.

2.3. Learning

It is assumed that learning in the brain is a consequence of changes in the synapses between neurons. These changes can happen in many different ways: the amount or type of neurotransmitter and receptors available, the surface area that connect both neurons, etc. By changing the way some neurons affect others it is possible to change the behavior of the network as a whole, adding, removing or modifying activity patterns in response to external or internal signals. Biologically plausible computational models composed of SNNs able to learn a target function have demonstrated being increasingly successful. Combining SNNs with the use of local learning rules, these models can be implemented in highly efficient, low-power, neuromorphic hardware (Rajendran et al., 2019). In computational models of neural networks a synapse has a quantifiable influence over the target

neuron (a weight). Learning in these models is implemented as the change over time of the network weights to make the system more capable to do some task.

In SNNs models a widely used learning rule is the spike-timing-dependent plasticity (STDP), a synaptic model with weight adaptation demonstrated in biological systems (Levy and Steward 1983) and more particularly in the BG (Fino and Venance 2010). In this rule, the weight of the synapse changes depending on the relative timing of presynaptic and postsynaptic spikes. If the spikes are far in time, there is no weight change, but the closer the spikes get in time (around tens of milliseconds or less), the stronger the weight change. The direction of the weight change might depend on the order of the spikes. Typically, if the presynaptic spike precedes (or follows) the postsynaptic spike then the weight increases (or decreases), as portrayed in Fig 3. (left). This does not always have to be the case though, and different timing relationships can occur.

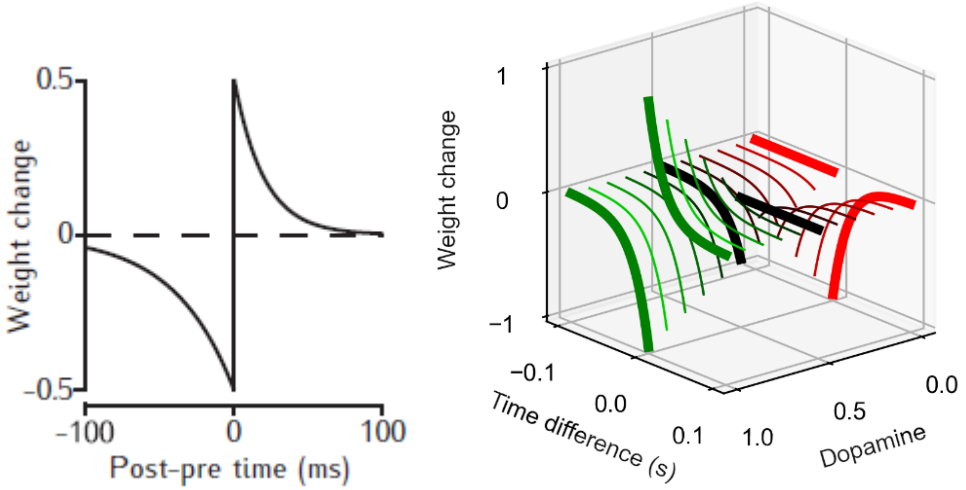

Figure 3**:** Different kernels used in spike-timing-dependent plasticity-like rules. (Left) Typical STDP kernel shape, showing the relationship between the relative spike timing and the weight change. (Right) An example of kernel used in STDE learning rule, where the weight change depends not only on the post-pre time difference but also on the available amount of dopamine.

A postsynaptic neuron equipped with STDP can detect and recognize the presence of repetitive patterns (Masquelier et al. 2009). This can be useful in unsupervised learning tasks, where data given without explicit target needs to be clustered. The STDP uses statistical correlations to strengthen synaptic connections, so what is learnt is biased to the most frequent patterns (Garrido et al. 2016). We instead want to bias the learning by the reward or punishment obtained, which in biological systems is signaled by the presence of extracellular DA. The STDP rule can be modified by adding the amount of DA as a factor of the weight change. This way is known as R-STDP and drives the learning of patterns that statistically correlate with a reward signal (Izhikevich 2007; Legenstein, Pecevski, and Maass 2008). As rewards tend to happen some time after the stimuli that caused the reward in the first place, it is needed to make somehow explicit this relationship between stimuli that are separated in time. This can be solved by the use of the so-called eligibility traces: variables that temporarily store the potential synaptic change until some amount of DA is received. The value stored decays exponentially over a span of seconds. In case a reward happens during this interval, a DA

signal is received in this synapse and the potential weight change stored in the eligibility trace is then applied. This way, rewards do not need to happen instantly after the relevant stimulus to be learnt, they can be delayed.

Neurons with synapses using R-STDP can learn to represent stimuli correlated with rewards, but learning in the striatum is a bit more complex. A more flexible synaptic model is proposed called Spike-Timing-Dependent Eligibility (STDE) based on physiological data that captures many features found in the biological MSN of the basal ganglia (K. N. Gurney, Humphries, and Redgrave 2015). This model is more flexible than the previous STDP-like rules as different learning kernels can be used depending on the amount and type (reward or punishment) of reinforcement received (Fig. 3, right). Although the authors did not include some important BG features like the GPe nucleus or a cortico-striatal loop, their model successfully learned to select an action channel driven by stronger cortical input, based only on the timing of the input and the reward signal.

DA is not the only neuromodulator influencing the striatum. ACh seems to have an important role regulating learning in MSNs of the STR, as ACh pauses define the time window for phasic dopamine to induce plasticity (Reynolds et al. 2022). However, it remains under discussion how these two mechanisms combine to make the striatum able to solve action-selection problems. An original proposal of this thesis is a computational model of the striatum that uses both DA and ACh to learn to map from stimulus to action, by using DA as a global reward signal that modulates the kernel of the STDP-like learning rule, and ACh as a local population feedback that signals the responsibility of the recent actions. With the ACh feedback the model can learn faster, even if the task to learn is more complex. In real brains, ACh is modulated at STR by thalamic inputs. Our model tests how this input can possibly play a role in facilitating learning by constraining the synaptic adaptation to specific subpopulations within STR.

Chapter 3: Results

1. Contributions to Specific Journals

Article #1: A Basal Ganglia Computational Model to Explain the Paradoxical Sensorial Improvement in the Presence of Huntington's Disease

González-Redondo Á, Naveros F, Ros E, Garrido JA. Int J Neural Syst. 2020 Oct;30(10):2050057. doi: 10.1142/S0129065720500574. Epub 2020 Aug 24. PMID: 32840409.

Relevance:

- Impact Factor (JCR 2020): 5.866
- Subject Category:
    - Computer Science, Artificial Intelligence. Ranking 28/139 (Q1)

Contribution summary:

The article introduces a computational model of the basal ganglia (BG) to investigate the apparent paradox of sensory improvement observed in patients with Huntington's disease (HD). This model encompasses the primary nuclei of the BG and examines the impact of altered dopamine levels and the influence of HD on information processing within the BG. The results indicate that the early and intermediate stages of HD may intensify transient activity in both the striatum and the substantia nigra pars reticulata (SNr), offering a potential explanation for the seemingly paradoxical enhancement in discrimination task performance. Moreover, the study delves into dopamine's role within the BG and its effect on the BG's performance as a selection mechanism. The findings propose that moderate levels of dopamine could enhance performance in selection tasks, while exceedingly high or low levels may prove detrimental. Overall, this study sheds light on the neural mechanisms responsible for the paradoxical sensory improvement in HD and promotes an improved understanding of the rapid dynamics present in the BG network.

Article #2: Black-box and surrogate optimization for tuning spiking neural models of striatum plasticity

Cruz NC, **González-Redondo Á**, Redondo JL, Garrido JA, Ortigosa EM and Ortigosa PM (2022). Front. Neuroinform. 16:1017222. doi: 10.3389/fninf.2022.1017222

Relevance:

- Impact Factor (JCR 2022): 3.5
- Subject Category:
    - Mathematical & Computational Biology. Ranking 15/55 (Q2)
    - Neurosciences. Ranking 126/272 (Q2)

Contribution summary:

The aim of this scientific article is to examine the challenge of adjusting spiking neural models of striatum plasticity through numerical optimization. The authors concentrate on a biologically inspired network model of the striatum, which captures significant experimental features and can recognize intricate input patterns. However, manual tuning of this model proves to be both difficult and time-consuming. The article explores the application of optimization algorithms to automate the tuning process and enhance the model's performance.

Four optimization methods are compared in this study: SurrogateOpt, RBFOpt, DIRECT-GL, and random search. These methods are tailored for black-box optimization and have minimal requirements for solution evaluations, rendering them suitable for computationally demanding models. The findings demonstrate that SurrogateOpt is the optimal choice for adjusting the spiking neural model, and the performance of the other methods is also discussed.

In conclusion, this study effectively showcases the use of optimization algorithms to automate the tuning of spiking neural models. It offers a suggestion for the most efficient approach to save time and improve the performance of these models. The implications of these findings extend to the development of computational models of the brain and the understanding of learning mechanisms.

Article #3: Reinforcement learning in a spiking neural model of striatum plasticity

González-Redondo, Á., Garrido, J., Arrabal, F. N., Kotaleski, J. H., Grillner, S., & Ros, E. (2023). Neurocomputing, 126377. doi: 10.1016/j.neucom.2023.126377

Relevance:

- Impact Factor (JCR 2022): 6.0
- Subject Category:
    - Computer Science, Artificial Intelligence. Ranking 41/145 (Q2)

Contribution summary:

The article introduces a computational model of the striatum, a part of the basal ganglia thought to be involved in action-selection based on reinforcement learning. The model incorporates several biologically inspired mechanisms, such as spike-timing-dependent plasticity (STDP), homeostatic mechanisms, and asymmetric lateral inhibitory connections. The authors show that their model can learn to choose the most rewarding actions in response to intricate input patterns.

Additionally, they explore the role of various neuronal and network features, including homeostatic mechanisms and lateral inhibitory connections, in action-selection. The findings indicate that homeostatic mechanisms render learning more robust and facilitate recovery following rewarding policy swaps, while lateral inhibitory connections play a significant role when multiple input patterns

are associated with the same rewarded action. The authors also discover that the optimal delay between the action and dopaminergic feedback is approximately 300ms, consistent with prior studies.

In summary, the model offers insights into the neural basis of decision-making, providing a biologically plausible explanation for the striatum's role in reinforcement learning.

2. Other Preliminary Results

The published works represent only a portion of the comprehensive research conducted and planned for the future. In particular, investigations have been conducted on models of Reward Prediction Error (RPE) and models that combine dopamine (DA) and acetylcholine (ACh) systems. It is worth noting that the ultimate goal is to integrate these findings into a cohesive, unified framework, which will be elaborated upon in the conclusions and future work sections of this research. In the following sections, we will provide a detailed explanation of the unpublished works that contribute to this broader research endeavor.

2.1. Striosome Model for Reward Prediction Error

Introduction

The basal ganglia are involved in various functions including motor control, emotions, and learning. The role of the basal ganglia in reinforcement learning is especially well-established, with dopaminergic neurons playing a key role in signaling reward prediction errors (Schultz, Dayan, and Montague 1997; Sutton and Barto 2018). The striatum, a primary component of the basal ganglia, is subdivided into two distinct compartments known as the striosomes (or patches) and the matrix (or matrisomes). Striosomes are small, densely packed clusters of neurons embedded within the matrix, and they are thought to play a significant role in reward processing and motivational aspects of behavior (Crittenden and Graybiel 2011; Fujiyama et al. 2011). They receive input from limbic areas and project to dopamine-producing regions such as the substantia nigra pars compacta (SNc) and ventral tegmental area (VTA), which are crucial in the computation of reward prediction errors and novelty detection (Horvitz 2000). Although there is a growing body of literature on basal ganglia models, computational models specifically targeting the striosomal compartments are relatively scarce [but see (Berthet et al. 2016; Amemori, Gibb, and Graybiel 2011; Shivkumar, Muralidharan, and Chakravarthy 2017)]. Understanding striosomal function is critical for advancing our knowledge of the neural substrates underlying reward-based learning and behavior.

In light of this, our study aims to contribute to the relatively unexplored field of striosomal computational models. Acknowledging the exploratory nature of this study, our primary focus was to develop a functional model that captures the essential aspects of striosomal operations in computing RPE. Specifically, we aim for the model to differentiate between predictable rewards/punishments and unexpected stimuli. Our hypothesis is that our computational model of striosomes will be capable of computing RPE by showing differentiated responses to expected rewards/punishments and

unexpected situations. While we endeavored to select biologically plausible parameters, achieving high biological realism was not the main objective at this stage. Future work may delve deeper into refining the model to better reflect the biological intricacies of striosomal function in reward processing.

Methods

Network Design

There are two major types of medium spiny neurons in the striatum: those expressing dopamine D1 receptors (D1-MSNs), which are generally thought to facilitate behavior ("Go" pathway), and those expressing D2 receptors (D2-MSNs), which are thought to inhibit behavior ("No-Go" pathway) (Gerfen 1992). Striosomes have been observed to have a relatively high concentration of D1-MSNs, and these D1-MSNs in striosomes project directly to the SNc, modulating dopaminergic signaling (Gerfen 1992). Our simulated network was designed based on these principles. The network topology consisted of an input layer with 2,000 neurons, a striosome layer with 80 neurons, and a dopaminergic neuron layer (SNc) with 1 neuron. The connections included input layer to striosome layer with spike-timing-dependent (STDE) synapses, striosome layer to SNc layer with STDE synapses and to itself with inhibitory synapses, and SNc layer to striosome layer and to itself with dopamine (DA) synapses, modulating both input to striosome and striosome to SNc connections (see Fig. 4 for a schematic representation of the network topology).

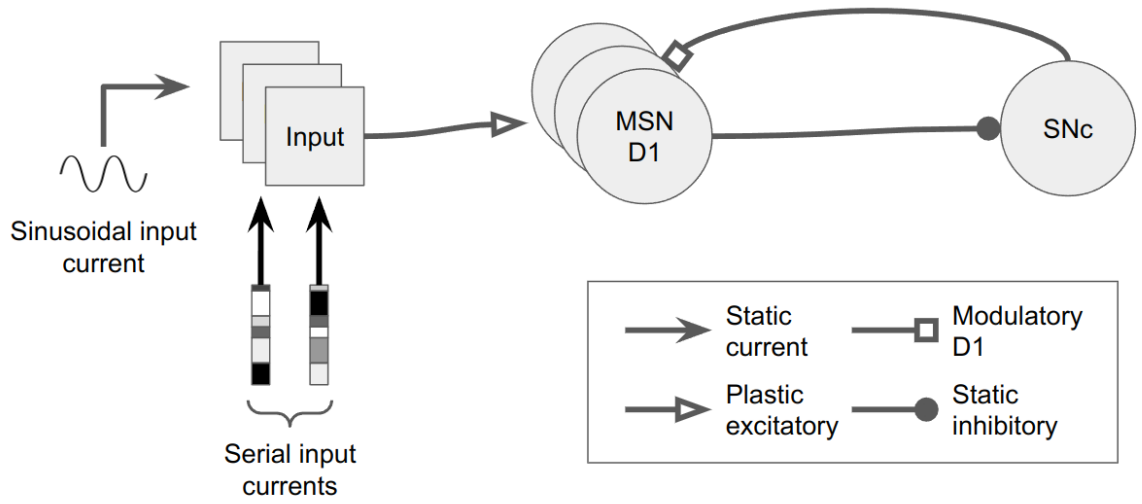

Figure 4**:** Striosomal network model for reward prediction error.

Models and Parameters

Leaky Integrate-and-Fire (LIF) model with time-driven dynamics was employed for input and striosome neurons, as in González-Redondo (2023). For the dopaminergic neurons in the SNc layer, the Izhikevich model was used, which offers a bit more flexibility in terms of capturing different neural

behaviors. This was done to keep options open for future experiments, where we might want to explore more complex dynamics without changing the neuron model. The used parameters are shown in the tables 1 and 2. All neurons began the simulation with a membrane potential equal to their leakage reversal potential or reset value. The synapse models included STDE for input to striosome synapses and striosome to SNc synapses. Static excitatory and inhibitory synapses were used for other connections. The learning rules were based on dopamine-based Spike-Timing-Dependent Plasticity (STDE) for input to striosome synapses ($STDE_{D1}$) and striosome to SNc synapses ($STDE_{SNC}$), as defined in González-Redondo (2023). Different kernel parameters were used for $STDE_{D1}$ and $STDE_{SNC}$ learning rules, as shown in table 3.

Parameter	Value	Description
`c_m`	50	Membrane capacitance (in pF)
`e_exc`	0	Excitatory reversal potential (in mV)
`e_inh`	-85	Inhibitory reversal potential (in mV)
`e_leak`	-65	Leakage reversal potential (in mV)
`g_leak`	10	Leakage conductance (in nS)
`tau_exc`	5	Time constant for excitatory synapses (in ms)
`tau_inh`	125	Time constant for inhibitory synapses (in ms)
`tau_nmda`	20	NMDA time constant (in ms)
`tau_ref`	1	Refractory time constant (in ms)
`v_thr`	-50	Threshold voltage for spike initiation (in mV)
`tau_thr`	50	Threshold time constant (in ms)
`tar_fir_rat`	0,4	Target firing rate

Table 1: LIF parameters used.

Parameter	Value	Description
`a`	0,1	Recovery variable time scale
`b`	0,2	Sensitivity of the recovery variable to subthreshold fluctuations
`c`	-65	After-spike reset value of the membrane potential (in mV)
`c_m`	10	Membrane capacitance (in pF)
`d`	2	After-spike reset of the recovery variable
`e_exc`	0	Excitatory reversal potential (in mV)
`e_inh`	-80	Inhibitory reversal potential (in mV)
`tau_exc`	1	Time constant for excitatory synapses (in ms)
`tau_inh`	62,5	Time constant for inhibitory synapses (in ms)
`tau_nmda`	20	NMDA time constant (in ms)

Table 2: Izhikevich parameters used.

		Value	
Parameter	Description	Input to STR	STR to SNc
`tau_plu`	Time constant for potentiation (s)	0,032	0,032
`tau_min`	Time constant for depression (s)	0,032	0,032
`tau_eli`	Eligibility trace time constant (s)	0,2	0,375
`tau_dop`	Dopamine time constant (s)	0,125	0,125
`inc_dop`	Increment factor for dopamine	8	8

`dop_max`	Maximum dopamine level	250	250
`dop_min`	Minimum dopamine level	50	50
`syn_pre_inc`	Synaptic increment factor	1,00E-04	0
`k_plu_hig`	High threshold potentiation factor	0,2	3,00E-03
`k_plu_low`	Low threshold potentiation factor	-1,00E-01	-3,00E-03
`k_min_hig`	High threshold depression factor	-0,2	3,00E-03
`k_min_low`	Low threshold depression factor	-1,00E-01	-3,00E-03

Table 3: Connectivity parameters used.

The simulation parameters included a time step of 1e-4 seconds and the Euler method for integration. All simulations were conducted using the EDLUT neural simulation engine (Carrillo et al. 2018) (version 2021) on a personal computer.

Experimental procedure

The experimental procedure simulated a total length of 200 seconds, divided into blocks of 1/8 seconds each, where each stimulus was presented. Each input pattern was randomly generated, with each component taken randomly from an uniform interval from 0 to maximum intensity (which generates at most 4 spikes per cycle). The input layer received sinusoidal current and the input patterns as currents. The SNc layer received a current input that varied based on reinforcement, plus the inhibitory input from the striosome layer.

To be able to test our hypothesis, the experimental procedure involved simulating the network's response to various input patterns and reinforcement values over time. The input patterns were fed into the input layer sequentially, while the reinforcement values modulated the input current to the SNc neuron. The simulation tracked the spiking activity of the neurons and the changes in synaptic weights throughout the experiment, allowing for the investigation of the network's learning dynamics and the interaction between the striosome and SNc layers in response to different input patterns and reinforcement values.

To evaluate the efficacy of our striosome model in computing RPE and to test our hypothesis that the model can differentiate between expected rewards/punishments and unexpected situations, we subjected the network to two distinct experimental conditions at the conclusion of the simulation. The first was the 'expected reward omission condition' (occurring in simulation blocks between 194 and 196 seconds), where rewards that the network had been conditioned to expect were omitted. This condition tested the model's ability to compute RPE by recognizing the deviation from expected rewards. The second was the 'new inputs condition' (occurring in simulation blocks between 198.25 and 198.75 seconds), where the network was exposed to novel input patterns it had not encountered before. This condition tested the model's responsiveness to unexpected situations by introducing novel stimuli. By observing how the network responds under these conditions, we can assess whether our striosome model demonstrates the essential aspects of striosomal operations in computing RPE as stated in our hypothesis.

Results

Fig. 5 depicts the spiking activity and evolution of synaptic weights in the network across different phases: the warming up phase (first 50 seconds), the learning phase (up to approximately 150 seconds), the learned phase (150-200 seconds), and the testing phase (final 10 seconds).

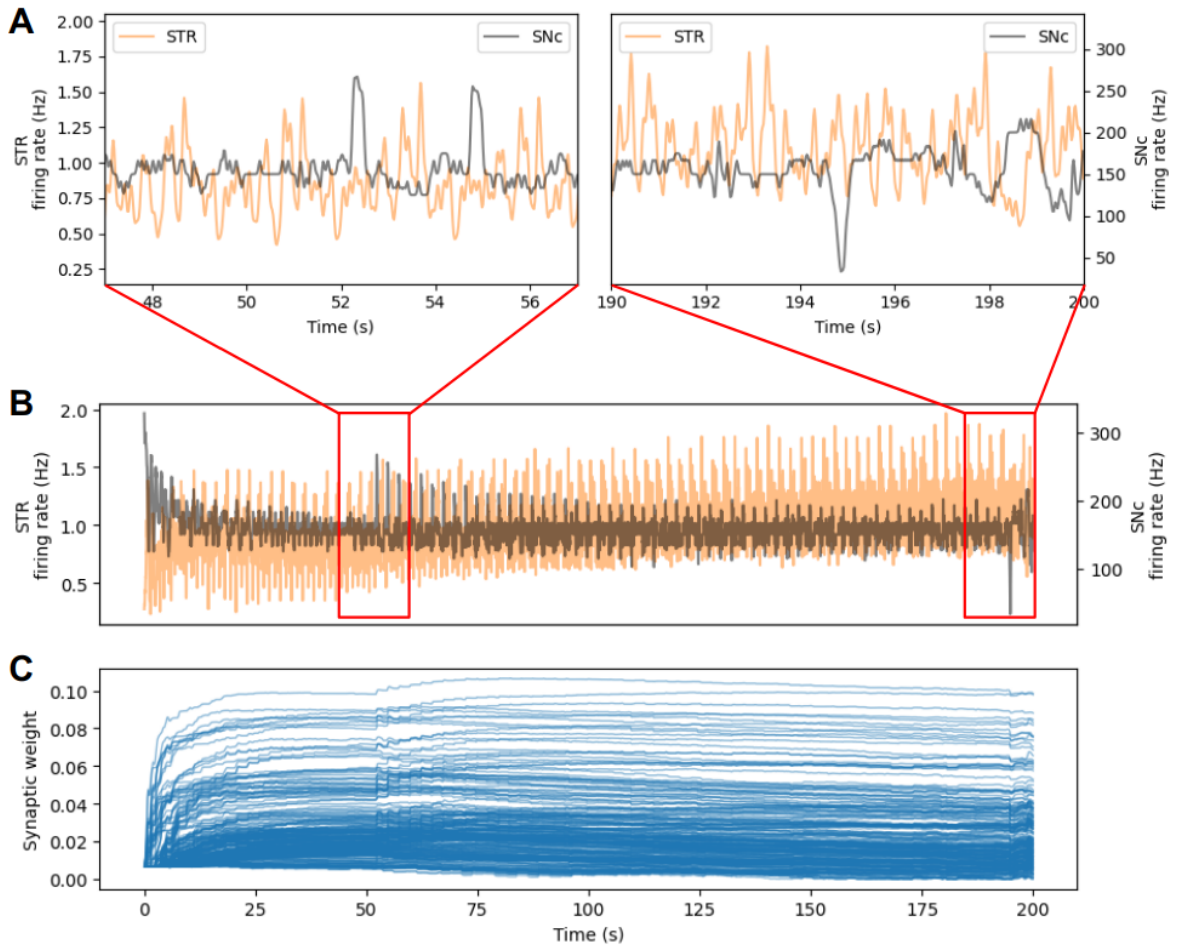

Figure 5**:** Network Activity (B) and Synaptic Weights Evolution (C). The insets (A) show in detail the activity during the beginning of the training phase (left) and at the testing phase (right, the last 10 seconds of simulation).

During the initial warming up phase, both striosomal (STR) and substantia nigra pars compacta (SNc) neurons exhibited elevated levels of activity. However, as the weights of the synapses from the STR to SNc increased, a rapid decay in the activity was observed. After approximately 25 seconds, the activity of the network stabilized, with the SNc neurons maintaining an activity level around 160 Hz and the STR neurons at approximately 1 Hz.

The learning phase started at the 50-second mark, with rewards being consistently delivered for the last two input patterns in the repeating sequence. During this period, the SNc neurons' activity showed peaks in response to the unexpected rewards, which were introduced at the beginning of the learning phase. Over the course of approximately 100 seconds, these peaks gradually reduced as the synapses from the input to STR and STR to SNc adapted, demonstrating the model’s capability to learn and predict these rewards.

During the learning phase, the network's response to rewards stabilized. To test the hypothesis that the model can detect unexpected situations, the simulation was then subjected to the expected reward omission condition starting at around 195 seconds. This led to a pronounced decrease in SNc activity, indicating the network's ability to detect the unexpected absence of rewards, in line with our hypothesis regarding the model's capacity for computing RPE in varying reward conditions.

In the final testing phase, novel input patterns that had never been encountered by the network were introduced at approximately 198 seconds under the new inputs condition. This was done to test the hypothesis regarding the model's ability to detect unexpected situations and compute RPE. This produced two cascading effects: initially, there was a decrease in the activity of STR neurons, followed by a surge in activity within the SNc neurons. The increase in SNc activity in response to novel input patterns demonstrates the model's capacity to react to novelty, supporting our hypothesis that it can effectively compute RPE in response to unexpected situations.

Discussion

The testing phase illustrates the network’s capability to adapt and respond to new circumstances. The initial dip in STR activity upon the introduction of novel input patterns suggests an inherent processing latency or processing adjustment within the striosomes. In contrast, the subsequent rise in SNc activity can be interpreted as the network's signal of a mismatch between predicted and actual outcomes, indicating its capacity to detect deviations from learned patterns.

The deep dip in SNc activity observed in response to the unexpected omission of rewards during the learned phase is consistent with the behavior expected from a model that computes reward prediction errors. This illustrates the network’s ability to generate RPE signals that correspond to the absence of anticipated rewards.

In conclusion, the results demonstrate that our striosome model is capable of learning from reward signals and effectively computing reward prediction errors, as evidenced by its responses to novel stimuli and the omission of expected rewards.

2.2. Dopamine and Acetylcholine Modulation in a Reinforcement Learning Striatal Model

Introduction

This study explores the role of DA and ACh in action-selection problem-solving within the striatum, a key component of the basal ganglia. We developed a reinforcement learning computational model of the striatum that leverages DA as a global reward signal and ACh as a local population feedback, leading to the mapping from stimulus to action. The model was enriched with lateral connectivity and homeostatic mechanisms, increasing its robustness to parametric changes. The model demonstrated proficiency in recognizing relevant patterns and consistently selecting rewarded actions, while its homeostatic mechanisms facilitated robust learning and recovery from policy changes. Notably, incorporating ACh feedback expedited the learning process as the number of potential actions increased. This study's findings provide a promising basis for future exploration into the intricate learning mechanisms of the brain and the role of neuromodulators therein.

The Role of Dopamine and Acetylcholine in the Striatal Model

DA and ACh are key neuromodulators in the basal ganglia's learning process. Phasic DA carries reinforcement signals to the STR, while ACh regulates learning in the MSNs of the STR by defining the window for phasic dopamine to induce plasticity (Reynolds et al. 2022). Despite their significant roles, it is still unclear how these two mechanisms work together to facilitate action-selection problem-solving in the striatum.

We hypothesize that ACh delineates a learning window, within which the global signal of DA can effectuate localized synaptic changes. This interplay between ACh and DA not only is instrumental in adapting action-selection based on reinforcement signals but also facilitates faster and more scalable learning by concentrating the learning process on the pertinent segments of the network.

To investigate this hypothesis, we constructed a reinforcement learning (RL) computational model of the striatum that is capable of learning the mapping from stimulus to action. The model integrates the following features:

- Dopamine functions as a global reward signal, modulating the kernel of the STDP learning rule.
- Acetylcholine serves as a local population feedback, signaling the relevance of recent actions.
- Additionally, we further enhanced our model by incorporating lateral connectivity and homeostatic mechanisms, thus enhancing the network's robustness to variations in parameters.

Methods

Network Model Structure

The network, modeled after (González-Redondo et al. 2023) and shown in Fig. 6, consists of Leaky Integrate-and-Fire (LIF) neurons organized into channels, with each representing a potential action. Each channel encompasses two MSN populations (D1 and D2 neurons), with lateral inhibition incorporated. Action neurons, which simplify other basal ganglia nuclei by integrating excitatory

activity from D1 neurons and inhibitory activity from D2 neurons, select an action if the activity balance between its D1 and D2 neurons leans toward D1, causing the corresponding action neuron to spike.

Role of the Environment and Feedback Mechanisms

The environment generates a 200-ms-delayed reinforcement signal based on the last action taken and the expected action. A dopaminergic neuron sends a global reward signal to all MSNs.

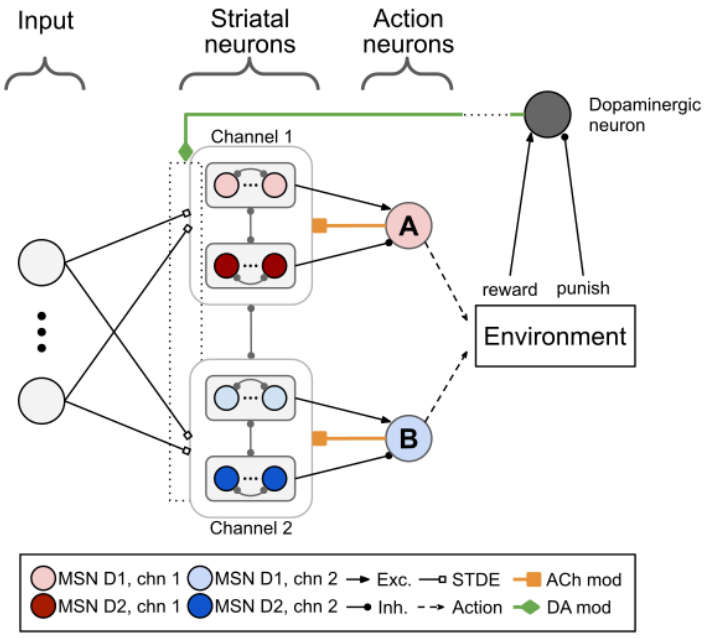

Figure 6**:** Structure of the cortico-striatal network solving a RL task.

Importantly, for this work, we added that action neurons also transmit information about the decision made back to the MSNs, and if the action was taken, ACh levels momentarily dip.

Learning rule

We use Spike-Timing Dependent Eligibility (STDE, (K. N. Gurney, Humphries, and Redgrave 2015)) learning rule similar to STDP, but with DA-dependent kernel constants (low, medium and high DA level kernels shown in Fig. 7). We extended this learning rule by adding ACh modulation, making learning only possible when the ACh level is low in a channel.

This rule uses eligibility traces that decay exponentially with a time constant of 400 ms to store the potential weight changes and apply them according to current DA level, similarly to (K. N. Gurney, Humphries, and Redgrave 2015; Izhikevich 2007). All plastic synapses share a global DA level that decays exponentially with a temporal constant of 20 ms.

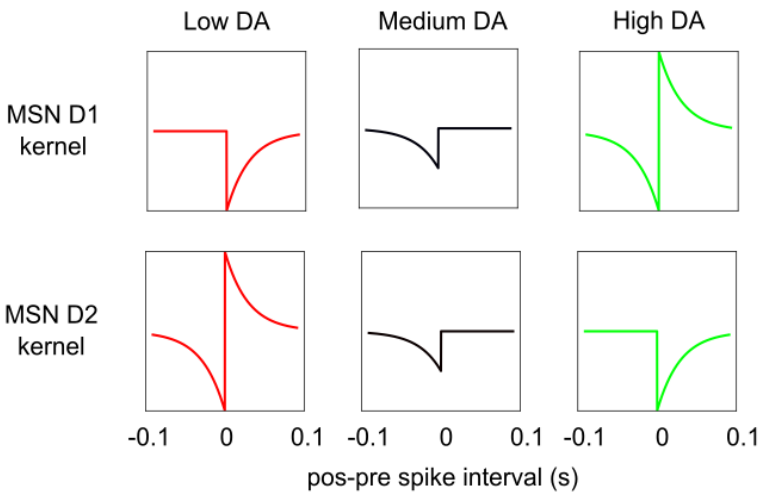

Figure 7: Different STDE learning kernels used by neuron type (D1 or D2) and by DA level (low or high).

Adaptive Threshold

To prevent neurons from becoming permanently silent during learning, we incorporated an adaptive threshold to the MSNs based on (Galindo et al. 2020). This modification makes neuron firing more sparse, balancing activity within the network. It also enhanced parameter suitability and its ability to recover swiftly after changes in rewarding policy, as already studied in González-Redondo et al. (2023) and shown in the next section.

Pattern Detection Task and Model Validation

Finally, to evaluate the robustness of our combined synaptic and homeostatic rules, we trained a single-neuron model to identify a specific pattern within a noisy input stream, as shown in Fig. 8. Two repeating patterns are presented 20% of the time each. If the striatal neuron fires in response to the rewarded pattern, reward is given. Conversely, if the striatal neuron fires in response to another pattern or noise, punishment is issued. In this simplified setting, ACh level remains consistently low.

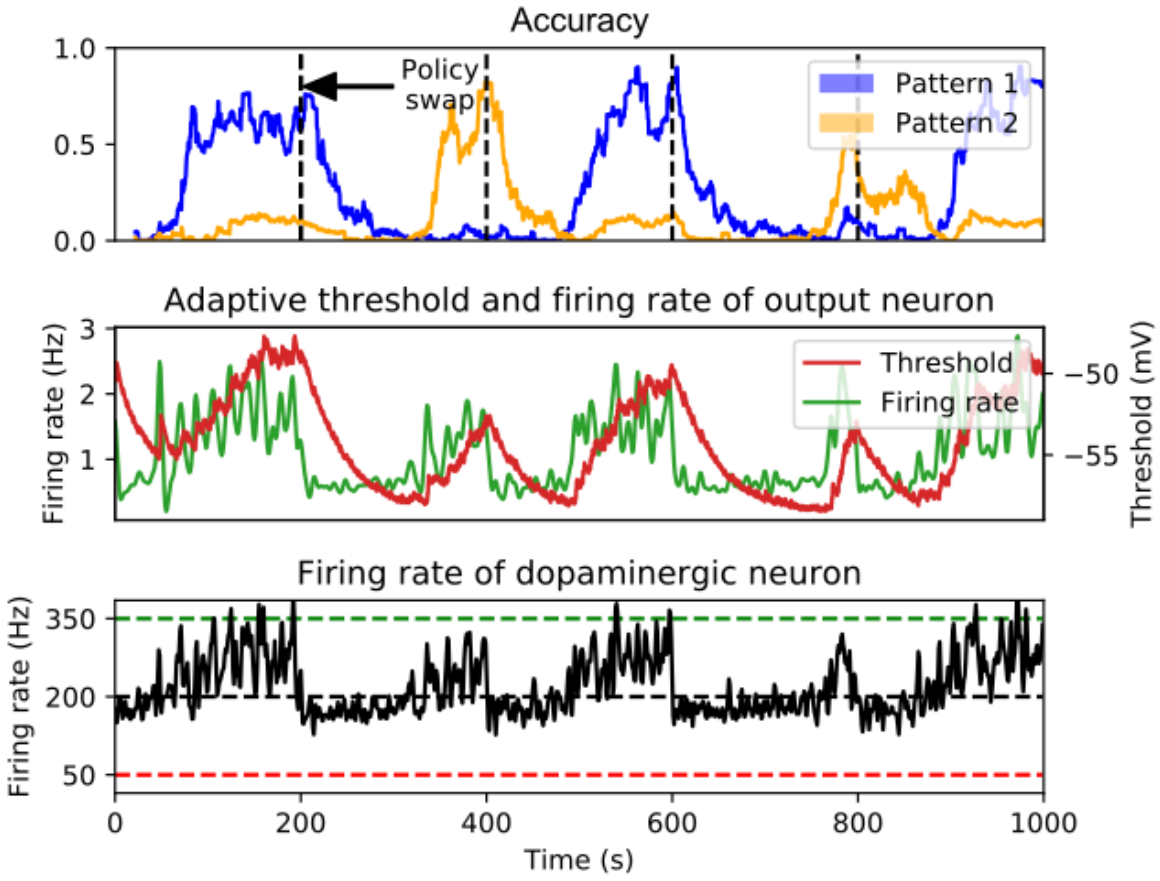

Figure 8**:** Pattern Detection Task Using a Single Neuron. Every 200 seconds, a distinct pattern (1 or 2) is associated with rewards when the neuron fires, facilitating learning (top row). The middle row displays the neuron's adaptive threshold (in red) and the firing rate (in green). The bottom row illustrates the firing rate of the dopaminergic neuron, which indicates the number of rewards the agent is receiving.

Results

The model's effectiveness was further tested through an action-selection experiment. In this scenario, the simulated network was presented with multiple possible actions to choose from. There are as many input patterns as possible actions, and these patterns were randomly shown during the simulation 80% of the time, with noise representing the remaining 20%. Each input pattern corresponded to a specific action that, if chosen, would yield a reward. Any other action would result in a punishment. If no action was taken, neither punishment nor reward was given.

This task was tested under various conditions, such as differing numbers of possible actions and with or without the presence of ACh. As the number of potential actions increased, the task's difficulty naturally escalated. This can be seen in Fig. 9, where the difficulty of the task increases with the number of possible actions, as it takes longer to achieve high accuracy. The graph also shows that with ACh the model learns much faster with a higher number of actions than the model without ACh. The only situation with no difference is where only 2 actions are used.

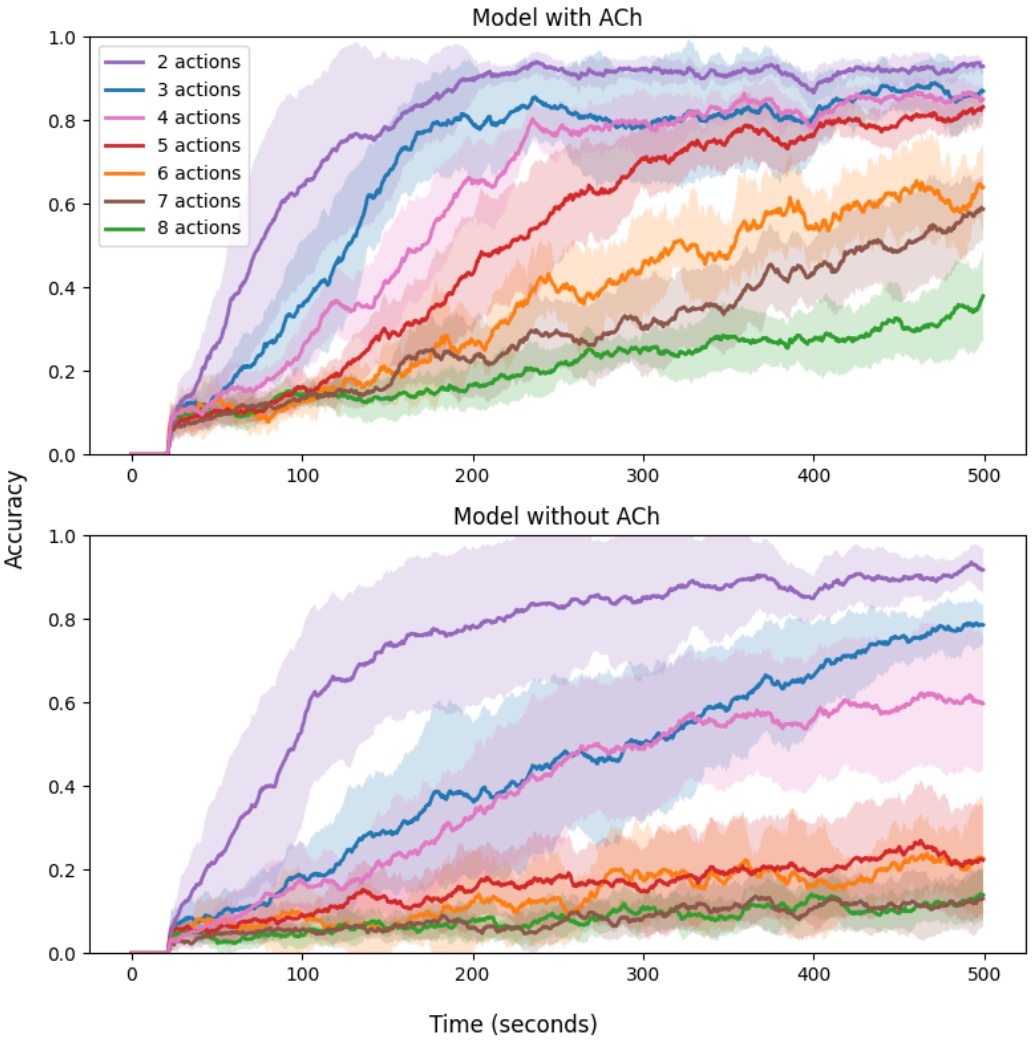

Figure 9: Comparison of the accuracy evolution achieved by a network with ACh (top) and without (bottom) when solving tasks of different difficulties (defined by the number of possible actions).

Discussion

The findings of this study suggest that the developed network model can effectively recognize relevant input patterns and make consistent, rewarding action choices in response to sensory inputs. An interesting observation was that the inclusion of ACh feedback expedited the learning process, particularly as the number of actions increased. In actual brains, ACh is modulated at STR by thalamic inputs. Our model suggests a potential role for this input in facilitating learning by confining it to specific subpopulations within the STR. This finding, along with others, offers an exciting path for further exploration in our understanding of the brain's complex learning mechanisms.

Chapter 4: Conclusions and Future Work

In this chapter we review the original objectives of the thesis to give context to the main scientific contributions done. Finally, we extract conclusions from the work done and propose future work based on our contributions.

1. Revisiting the Thesis Objectives

The primary objective of the thesis on computational neuroscience is to investigate the complex neural mechanisms and operations underlying decision-making, learning, and action selection, with a focus on the basal ganglia and its essential components. To fulfill this main objective, the following specific objectives were established and addressed in the thesis:

1. **Investigate basal ganglia-associated neurological disorders**, including Huntington's disease, using computational models. The published work (González-Redondo et al. 2020) presents a computational model that explores the unusual sensory improvements in patients with Huntington's disease. The study provides insights into the impact of dopamine level changes and disease-related factors on information processing and basal ganglia dynamics.
2. **Study action selection and reinforcement learning tasks** using basal ganglia models. This objective is addressed first in (González-Redondo et al. 2023), and then in the unpublished work explained in Chapter 2, section *2.2. DA and ACh Modulation in a RL Striatal Model*. In (González-Redondo et al. 2023) a computational striatum model is developed. It integrates biologically inspired mechanisms, such as STDP, homeostatic mechanisms, and lateral inhibitory connections, and demonstrates a capability for learning and selecting rewarding actions. The section *2.2. DA and ACh Modulation in a RL Striatal Model* examines the roles of neuromodulators DA and ACh in action selection, using a reinforcement learning computational model, thus improving our understanding of stimulus-to-action mapping.
3. **Examine the interaction between the basal ganglia model and other brain areas**, as well as their influence on motor learning during decision-making tasks. Work done in (Cruz et al. 2022) and in Chapter 2, section *2.1. Striosome Model for Reward Prediction Error* contributes to this objective. The article (Cruz et al. 2022) investigates the use of optimization algorithms in automating the tuning of spiking neural models to enhance computational models of the brain and facilitate interactions with other brain areas. Section *2.1. Striosome Model for Reward Prediction Error* examines the role of striosomes in reward processing, shedding light on the interactions between various brain structures and their impact on decision-making.

In conclusion, the specific objectives outlined for the thesis remain closely aligned with its overall purpose. This research contributes to a comprehensive understanding of basal ganglia functionality, the influence of neuromodulators, and the relationships between these elements and the functioning of the brain in normal and pathological conditions. While acknowledging potential limitations, the findings of this thesis have meaningful implications. On one hand, they can aid the development of

new approaches for understanding and treating basal ganglia-related neurological disorders. On the other hand, they can help in the creation of bioinspired control systems for embodied agents.

2. Main Contributions

1. A computational model of the basal ganglia to explain the paradoxical sensorial improvement observed in patients with Huntington's disease (González-Redondo et al. 2020):
    - The model included the main nuclei of the BG and simulated the effect of altered levels of dopamine and HD affectation on information processing in the BG.
    - The results showed that early and medium stages of HD affectation may enhance transient activity in the striatum and the substantia nigra pars reticulata (SNr), providing a possible explanation for the paradoxical improvement observed in discrimination task performance.
    - The findings suggested that medium levels of dopamine can improve performance in selection tasks, while high or low levels of dopamine may have detrimental effects.
2. An in-depth study of tuning spiking neural models of striatum plasticity (Cruz et al. 2022):
    - The study compared four optimization methods: SurrogateOpt, RBFOpt, DIRECT-GL, and random search.
        i. SurrogateOpt was found to be the most effective option for tuning the spiking neural model.
        ii. RBFOpt and random search yielded reasonable results, but were less effective than SurrogateOpt.
        iii. DIRECT-GL was the least effective among the methods tested.
    - The findings provide insights for the development of computational models of the brain and understanding learning mechanisms. Moreover, the study illustrates the utility of automatic optimization algorithms for facilitating more objective comparisons between models with different computational features, which can contribute to the robustness of the results and reduce reliance on manual tuning expertise.
3. Development of a functional and biologically inspired network model of the striatum (González-Redondo et al. 2023):
    - The model successfully integrated multiple features and mechanisms, and demonstration that the proposed model can learn to recognize complex input patterns and consistently choose rewarded actions in response to those patterns.
    - Analysis of the role of homeostatic mechanisms in making learning more robust and facilitating recovery after rewarding policy swapping.
    - Investigation of the importance of lateral inhibitory connections when multiple input patterns are associated with the same rewarded action.
    - Use of a spiking neural network with spike-time pattern representation that scales well with different pattern complexities, making the model suitable for a wide range of reinforcement learning tasks.
    - Finding that the optimal delay between the action and the dopaminergic feedback is

around 300ms, consistent with previous studies.

4. Developed a functional striosome model for reward prediction error (RPE) in the basal ganglia (Chapter 3, Section *2.1. Striosome Model for Reward Prediction Error*).
    - Designed a network topology consisting of an input layer, a striosome layer, and a dopaminergic neuron layer (SNc).
    - Demonstrated the model's capability to learn from reward signals and effectively compute RPE.
    - Showed the network's adaptability and capacity to detect deviations from learned patterns in response to novel stimuli and the omission of expected rewards.
5. Refinement of the previous striatum RL model by incorporating acetylcholine (ACh) as a local population feedback (Chapter 3, Section *2.2. Dopamine and Acetylcholine Modulation in a Reinforcement Learning Striatal Model*).
    - Developed a reinforcement learning computational model of the striatum that leverages dopamine (DA) as a global reward signal and acetylcholine (ACh) as a local population feedback.
    - Demonstrated model's proficiency in recognizing relevant patterns and consistently selecting rewarded actions.
    - Found that incorporating ACh feedback expedited the learning process as the number of potential actions increased.

3. Conclusions and Future Work

Our work has focused on the investigation and modeling of the neural mechanisms underlying action-selection, learning, and decision-making processes, with emphasis on the striatum and the role of dopamine, acetylcholine, and plasticity within the basal ganglia. Through computational modeling and innovative methodologies, the research shown collectively enhances our understanding of the complex interactions between neuronal populations and neuromodulators in the basal ganglia. These findings provide a promising foundation for future research into the brain's intricate learning mechanisms. As we stated, this can have practical applications for both improving the understanding and treating neurological disorders such as Huntington's disease, and also the development of bioinspired reinforcement agents. However, we acknowledge limitations and call for further research to validate and expand upon our results, incorporating additional biological factors and exploring the potential of optimization techniques and high-performance computing platforms in refining neural models.

More specifically, our findings demonstrate the potential for computational modeling and innovative methodologies to uncover the intricate workings of the brain's reinforcement learning systems. For instance, we present biologically inspired network models that incorporate striatal spike-timing-dependent plasticity, lateral connectivity, and homeostatic mechanisms. The incorporation of ACh as a second modulator for accountability signaling is, as far as we know, an original proposal with biological and computational support. These models have been shown to

significantly enhance the learning and re-learning of rewarded patterns, emphasizing the potential of these mechanisms to improve reinforcement learning models, and ultimately, our understanding of the brain's complex decision-making processes.

Moreover, our research has elucidated ways in which neural models can be optimized, such as through the use of SurrogateOpt, an optimization-based technique for tuning spiking neural models of striatum plasticity. The application of these techniques has been shown to yield more reliable and accurate learning capabilities when compared to manually tuned results. As such, these findings underscore the potential of optimization techniques in advancing our comprehension of brain reinforcement learning mechanisms, which are critical for motor control and decision-making tasks.

In addition to providing valuable insights into the normal functioning of neural networks within the brain, our research offers potential implications for understanding and treating neurological disorders such as Huntington's disease. For example, our first study presents a computational model of the basal ganglia that explains the paradoxical improvement in sensorial discrimination observed in Huntington's disease patients. Insights gained from these models may prove instrumental in guiding future research and developing novel therapeutic interventions for such conditions.

However, it is essential to acknowledge the limitations present within our studies and the need for further research to validate and expand upon these findings. One significant limitation is that the internal dynamics of the models are sensitive to the form of input representation; specifically, we used a phase-of-firing encoding scheme where input neurons encode the intensity of their input as a specific phase within a sinusoidal wave. This encoding scheme is critical to the performance of the model, and variations in input encoding could affect results. Additionally, incorporating biological factors such as the recurrent loop between the basal ganglia and the cortex, phasic dopamine signals, and the role of interneurons in the striatum, will be important for a more comprehensive understanding of the involved neural mechanisms. Furthermore, the exploration of high-performance computing platforms and the development of novel optimization techniques will be essential in refining these neural models, contributing to more accurate representations of the brain's complex learning and decision-making processes.

Bibliography

Amemori, Ken-ichi, Leif Gibb, and Ann Graybiel. 2011. “Shifting Responsibly: The Importance of Striatal Modularity to Reinforcement Learning in Uncertain Environments.” *Frontiers in Human Neuroscience* 5. https://www.frontiersin.org/articles/10.3389/fnhum.2011.00047.

Antonietti, Alberto, Jessica Monaco, Egidio D’Angelo, Alessandra Pedrocchi, and Claudia Casellato. 2018. “Dynamic Redistribution of Plasticity in a Cerebellar Spiking Neural Network Reproducing an Associative Learning Task Perturbed by TMS.” *International Journal of Neural Systems* 28 (09): 1850020. https://doi.org/10.1142/S012906571850020X.

Antunes, Gabriela, Samuel F. Faria da Silva, and Fabio M. Simoes de Souza. 2018. “Mirror Neurons Modeled Through Spike-Timing-Dependent Plasticity Are Affected by Channelopathies Associated with Autism Spectrum Disorder.” *International Journal of Neural Systems* 28 (05): 1750058. https://doi.org/10.1142/S0129065717500587.

Berke, Joshua D. 2018. “What Does Dopamine Mean?” *Nature Neuroscience* 21 (6): 787–93. https://doi.org/10.1038/s41593-018-0152-y.

Berthet, Pierre, Mikael Lindahl, Philip J. Tully, Jeanette Hellgren-Kotaleski, and Anders Lansner. 2016. “Functional Relevance of Different Basal Ganglia Pathways Investigated in a Spiking Model with Reward Dependent Plasticity.” *Frontiers in Neural Circuits* 10. https://www.frontiersin.org/articles/10.3389/fncir.2016.00053.

Bogacz, Rafal. 2020. “Dopamine Role in Learning and Action Inference.” *ELife* 9: e53262. https://doi.org/10.7554/eLife.53262.

Brainard, Michael S., and Allison J. Doupe. 2002. “What Songbirds Teach Us about Learning.” *Nature* 417 (6886): 351–58. https://doi.org/10.1038/417351a.

Burke, Dennis A., Horacio G. Rotstein, and Veronica A. Alvarez. 2017. “Striatal Local Circuitry: A New Framework for Lateral Inhibition.” *Neuron* 96 (2): 267–84. https://doi.org/10.1016/j.neuron.2017.09.019.

Carrillo, Richard R., J. Garrido, F. Naveros, and N. R. Luque. 2018. “EDLUT Simulator.” Personal Web Site. March 25, 2018. https://www.ugr.es/~rcarrillo/post/edlut/.

Chesselet, Marie-Françoise, and Jill M Delfs. 1996. “Basal Ganglia and Movement Disorders: An Update.” *Trends in Neurosciences* 19 (10): 417–22. https://doi.org/10.1016/0166-2236(96)10052-7.

Cox, Julia, and Ilana B. Witten. 2019. “Striatal Circuits for Reward Learning and Decision-Making.” *Nature Reviews Neuroscience* 20 (8): 482–94. https://doi.org/10.1038/s41583-019-0189-2.

Crittenden, Jill, and Ann Graybiel. 2011. “Basal Ganglia Disorders Associated with Imbalances in the Striatal Striosome and Matrix Compartments.” *Frontiers in Neuroanatomy* 5. https://www.frontiersin.org/articles/10.3389/fnana.2011.00059.

Cruz, Nicolás C., Álvaro González-Redondo, Juana L. Redondo, Jesús A. Garrido, Eva M. Ortigosa, and Pilar M. Ortigosa. 2022. “Black-Box and Surrogate Optimization for Tuning Spiking Neural Models of Striatum Plasticity.” *Frontiers in Neuroinformatics* 16. https://www.frontiersin.org/articles/10.3389/fninf.2022.1017222.

Dahl, George E., Dong Yu, Li Deng, and Alex Acero. 2012. “Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition.” *IEEE Transactions on Audio, Speech, and Language Processing* 20 (1): 30–42. https://doi.org/10.1109/TASL.2011.2134090.

DeLong, M. R., M. D. Crutcher, and A. P. Georgopoulos. 1985. “Primate Globus Pallidus and Subthalamic Nucleus: Functional Organization.” *Journal of Neurophysiology* 53 (2): 530–43. https://doi.org/10.1152/jn.1985.53.2.530.

Dulac-Arnold, Gabriel, Daniel Mankowitz, and Todd Hester. 2019. “Challenges of Real-World Reinforcement Learning.” arXiv. https://doi.org/10.48550/arXiv.1904.12901.

Fino, Elodie, and Laurent Venance. 2010. “Spike-Timing Dependent Plasticity in the Striatum.” *Frontiers in Synaptic Neuroscience* 2. https://www.frontiersin.org/articles/10.3389/fnsyn.2010.00006.

Foster, Nicholas N., Joshua Barry, Laura Korobkova, Luis Garcia, Lei Gao, Marlene Becerra, Yasmine Sherafat, et al. 2021. “The Mouse Cortico–Basal Ganglia–Thalamic Network.” *Nature* 598 (7879): 188–94. https://doi.org/10.1038/s41586-021-03993-3.

Fujiyama, Fumino, Jaerin Sohn, Takashi Nakano, Takahiro Furuta, Kouichi C. Nakamura, Wakoto Matsuda, and Takeshi Kaneko. 2011. “Exclusive and Common Targets of Neostriatofugal Projections of Rat Striosome Neurons: A Single Neuron-Tracing Study Using a Viral Vector.” *European Journal of Neuroscience* 33 (4): 668–77.

https://doi.org/10.1111/j.1460-9568.2010.07564.x.
Galindo, Sergio E., Pablo Toharia, Óscar D. Robles, Eduardo Ros, Luis Pastor, and Jesús A. Garrido. 2020. “Simulation, Visualization and Analysis Tools for Pattern Recognition Assessment with Spiking Neuronal Networks.” *Neurocomputing* 400 (August): 309–21. https://doi.org/10.1016/j.neucom.2020.02.114.
Garrido, Jesús A., Niceto R. Luque, Silvia Tolu, and Egidio D’Angelo. 2016. “Oscillation-Driven Spike-Timing Dependent Plasticity Allows Multiple Overlapping Pattern Recognition in Inhibitory Interneuron Networks.” *International Journal of Neural Systems* 26 (05): 1650020. https://doi.org/10.1142/S0129065716500209.
Geminiani, Alice, Claudia Casellato, Alberto Antonietti, Egidio D’Angelo, and Alessandra Pedrocchi. 2018. “A Multiple-Plasticity Spiking Neural Network Embedded in a Closed-Loop Control System to Model Cerebellar Pathologies.” *International Journal of Neural Systems* 28 (05): 1750017. https://doi.org/10.1142/S0129065717500174.
Gerfen, C R. 1992. “The Neostriatal Mosaic: Multiple Levels of Compartmental Organization in the Basal Ganglia.” *Annual Review of Neuroscience* 15 (1): 285–320. https://doi.org/10.1146/annurev.ne.15.030192.001441.
Ghosh-Dastidar, Samanwoy, and Hojjat Adeli. 2007. “Improved Spiking Neural Networks for EEG Classification and Epilepsy and Seizure Detection.” *Integrated Computer-Aided Engineering* 14 (3): 187–212. https://doi.org/10.3233/ICA-2007-14301.
———. 2009a. “Spiking Neural Networks.” *International Journal of Neural Systems* 19 (04): 295–308. https://doi.org/10.1142/S0129065709002002.
———. 2009b. “A New Supervised Learning Algorithm for Multiple Spiking Neural Networks with Application in Epilepsy and Seizure Detection.” *Neural Networks* 22 (10): 1419–31. https://doi.org/10.1016/j.neunet.2009.04.003.
González-Redondo, Álvaro, Jesús Garrido, Francisco Naveros Arrabal, Jeanette Hellgren Kotaleski, Sten Grillner, and Eduardo Ros. 2023. “Reinforcement Learning in a Spiking Neural Model of Striatum Plasticity.” *Neurocomputing* 548 (September): 126377. https://doi.org/10.1016/j.neucom.2023.126377.
González-Redondo, Álvaro, Francisco Naveros, Eduardo Ros, and Jesús A. Garrido. 2020. “A Basal Ganglia Computational Model to Explain the Paradoxical Sensorial Improvement in the Presence of Huntington’s Disease.” *International Journal of Neural Systems* 30 (10): 2050057. https://doi.org/10.1142/S0129065720500574.
Graybiel, Ann M. 1998. “The Basal Ganglia and Chunking of Action Repertoires.” *Neurobiology of Learning and Memory* 70 (1): 119–36. https://doi.org/10.1006/nlme.1998.3843.
Grillner, Sten, Jeanette Hellgren, Ariane Ménard, Kazuya Saitoh, and Martin A. Wikström. 2005. “Mechanisms for Selection of Basic Motor Programs – Roles for the Striatum and Pallidum.” *Trends in Neurosciences* 28 (7): 364–70. https://doi.org/10.1016/j.tins.2005.05.004.
Gurney, K., T. J. Prescott, and P. Redgrave. 2001. “A Computational Model of Action Selection in the Basal Ganglia. I. A New Functional Anatomy.” *Biological Cybernetics* 84 (6): 401–10. https://doi.org/10.1007/PL00007984.
Gurney, Kevin N., Mark D. Humphries, and Peter Redgrave. 2015. “A New Framework for Cortico-Striatal Plasticity: Behavioural Theory Meets In Vitro Data at the Reinforcement-Action Interface.” *PLOS Biology* 13 (1): e1002034. https://doi.org/10.1371/journal.pbio.1002034.
Harrington, Deborah L., Kathleen Y. Haaland, and Neal Hermanowitz. 1998. “Temporal Processing in the Basal Ganglia.” *Neuropsychology* 12: 3–12. https://doi.org/10.1037/0894-4105.12.1.3.
Hart, Andrew S., Robb B. Rutledge, Paul W. Glimcher, and Paul E. M. Phillips. 2014. “Phasic Dopamine Release in the Rat Nucleus Accumbens Symmetrically Encodes a Reward Prediction Error Term.” *Journal of Neuroscience* 34 (3): 698–704. https://doi.org/10.1523/JNEUROSCI.2489-13.2014.
Hikosaka, Okihide, Yoriko Takikawa, and Reiko Kawagoe. 2000. “Role of the Basal Ganglia in the Control of Purposive Saccadic Eye Movements.” *Physiological Reviews* 80 (3): 953–78. https://doi.org/10.1152/physrev.2000.80.3.953.
Horvitz, J. C. 2000. “Mesolimbocortical and Nigrostriatal Dopamine Responses to Salient Non-Reward Events.” *Neuroscience* 96 (4): 651–56. https://doi.org/10.1016/S0306-4522(00)00019-1.
Hunnicutt, Barbara J, Bart C Jongbloets, William T Birdsong, Katrina J Gertz, Haining Zhong, and Tianyi Mao. 2016. “A Comprehensive Excitatory Input Map of the Striatum Reveals Novel Functional Organization.” Edited by David C Van Essen. *ELife* 5 (November): e19103. https://doi.org/10.7554/eLife.19103.
Izhikevich, Eugene M. 2007. “Solving the Distal Reward Problem through Linkage of STDP and

Dopamine Signaling." *Cerebral Cortex* 17 (10): 2443–52. https://doi.org/10.1093/cercor/bhl152.
Jankovic, J. 2008. "Parkinson's Disease: Clinical Features and Diagnosis." *Journal of Neurology, Neurosurgery & Psychiatry* 79 (4): 368–76. https://doi.org/10.1136/jnnp.2007.131045.
Krichmar, Jeffrey L. 2018. "Neurorobotics—A Thriving Community and a Promising Pathway Toward Intelligent Cognitive Robots." *Frontiers in Neurorobotics* 12. https://www.frontiersin.org/articles/10.3389/fnbot.2018.00042.
Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2017. "ImageNet Classification with Deep Convolutional Neural Networks." *Communications of the ACM* 60 (6): 84–90. https://doi.org/10.1145/3065386.
Legenstein, Robert, Dejan Pecevski, and Wolfgang Maass. 2008. "A Learning Theory for Reward-Modulated Spike-Timing-Dependent Plasticity with Application to Biofeedback." *PLOS Computational Biology* 4 (10): e1000180. https://doi.org/10.1371/journal.pcbi.1000180.
Levy, W. B., and O. Steward. 1983. "Temporal Contiguity Requirements for Long-Term Associative Potentiation/Depression in the Hippocampus." *Neuroscience* 8 (4): 791–97. https://doi.org/10.1016/0306-4522(83)90010-6.
Maass, Wolfgang. 1997. "Networks of Spiking Neurons: The Third Generation of Neural Network Models." *Neural Networks* 10 (9): 1659–71. https://doi.org/10.1016/S0893-6080(97)00011-7.
Maia, Tiago V., Rebecca E. Cooney, and Bradley S. Peterson. 2008. "The Neural Bases of Obsessive-Compulsive Disorder in Children and Adults." *Development and Psychopathology* 20 (4): 1251–83. https://doi.org/10.1017/S0954579408000606.
Masquelier, Timothée, Etienne Hugues, Gustavo Deco, and Simon J. Thorpe. 2009. "Oscillations, Phase-of-Firing Coding, and Spike Timing-Dependent Plasticity: An Efficient Learning Scheme." *Journal of Neuroscience* 29 (43): 13484–93. https://doi.org/10.1523/JNEUROSCI.2207-09.2009.
Mogenson, Gordon J., Douglas L. Jones, and Chi Yiu Yim. 1980. "From Motivation to Action: Functional Interface between the Limbic System and the Motor System." *Progress in Neurobiology* 14 (2): 69–97. https://doi.org/10.1016/0301-0082(80)90018-0.
Obeso, José A., C. Warren Olanow, and John G. Nutt. 2000. "Basal Ganglia, Parkinson's Disease and Levedopa Therapy." *Trends in Neurosciences* 23 (October): S1. https://doi.org/10.1016/S1471-1931(00)00060-4.
Parent, André, and Lili-Naz Hazrati. 1995. "Functional Anatomy of the Basal Ganglia. II. The Place of Subthalamic Nucleus and External Pallidium in Basal Ganglia Circuitry." *Brain Research Reviews* 20 (1): 128–54. https://doi.org/10.1016/0165-0173(94)00008-D.
Parés-Badell, Oleguer, Gabriela Barbaglia, Petra Jerinic, Anders Gustavsson, Luis Salvador-Carulla, and Jordi Alonso. 2014. "Cost of Disorders of the Brain in Spain." *PLOS ONE* 9 (8): e105471. https://doi.org/10.1371/journal.pone.0105471.
Redgrave, P., T. J. Prescott, and K. Gurney. 1999. "The Basal Ganglia: A Vertebrate Solution to the Selection Problem?" *Neuroscience* 89 (4): 1009–23.
Redgrave, Peter, Manuel Rodriguez, Yoland Smith, Maria C. Rodriguez-Oroz, Stephane Lehericy, Hagai Bergman, Yves Agid, Mahlon R. DeLong, and Jose A. Obeso. 2010. "Goal-Directed and Habitual Control in the Basal Ganglia: Implications for Parkinson's Disease." *Nature Reviews Neuroscience* 11 (11): 760–72. https://doi.org/10.1038/nrn2915.
Reynolds, John N. J., Riccardo Avvisati, Paul D. Dodson, Simon D. Fisher, Manfred J. Oswald, Jeffery R. Wickens, and Yan-Feng Zhang. 2022. "Coincidence of Cholinergic Pauses, Dopaminergic Activation and Depolarisation of Spiny Projection Neurons Drives Synaptic Plasticity in the Striatum." *Nature Communications* 13 (1): 1296. https://doi.org/10.1038/s41467-022-28950-0.
Reynolds, John N. J, and Jeffery R Wickens. 2002. "Dopamine-Dependent Plasticity of Corticostriatal Synapses." *Neural Networks* 15 (4): 507–21. https://doi.org/10.1016/S0893-6080(02)00045-X.
Schultz, Wolfram, Peter Dayan, and P. Read Montague. 1997. "A Neural Substrate of Prediction and Reward." *Science* 275 (5306): 1593–99. https://doi.org/10.1126/science.275.5306.1593.
Shipp, Stewart. 2017. "The Functional Logic of Corticostriatal Connections." *Brain Structure and Function* 222 (2): 669–706. https://doi.org/10.1007/s00429-016-1250-9.
Shivkumar, Sabyasachi, Vignesh Muralidharan, and V. Srinivasa Chakravarthy. 2017. "A Biologically Plausible Architecture of the Striatum to Solve Context-Dependent Reinforcement Learning Tasks." *Frontiers in Neural Circuits* 11. https://www.frontiersin.org/articles/10.3389/fncir.2017.00045.
Suárez, Laura E., Ross D. Markello, Richard F. Betzel, and Bratislav Misic. 2020. "Linking Structure and Function in Macroscale Brain Networks." *Trends in Cognitive Sciences* 24 (4): 302–15. https://doi.org/10.1016/j.tics.2020.01.008.

Suryanarayana, Shreyas M., Jeanette Hellgren Kotaleski, Sten Grillner, and Kevin N. Gurney. 2019. “Roles for Globus Pallidus Externa Revealed in a Computational Model of Action Selection in the Basal Ganglia.” *Neural Networks* 109 (January): 113–36. https://doi.org/10.1016/j.neunet.2018.10.003.

Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. Second edition. Adaptive Computation and Machine Learning Series. Cambridge, Massachusetts: The MIT Press.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. “Attention Is All You Need.” arXiv. https://doi.org/10.48550/arXiv.1706.03762.

Wolf, Steven S., Douglas W. Jones, Michael B. Knable, Julia G. Gorey, Kan Sam Lee, Thomas M. Hyde, Richard Coppola, and Daniel R. Weinberger. 1996. “Tourette Syndrome: Prediction of Phenotypic Variation in Monozygotic Twins by Caudate Nucleus D2 Receptor Binding.” *Science* 273 (5279): 1225–27. https://doi.org/10.1126/science.273.5279.1225.

Xiao, Lei, and Todd F. Roberts. 2021. “What Is the Role of Thalamostriatal Circuits in Learning Vocal Sequences?” *Frontiers in Neural Circuits* 15. https://www.frontiersin.org/articles/10.3389/fncir.2021.724858.

Annex: Journal Articles Included in this Thesis

International Journal of Neural Systems, Vol. 30, No. 10 (2020) 1–18

A BASAL GANGLIA COMPUTATIONAL MODEL TO EXPLAIN THE PARADOXICAL SENSORIAL IMPROVEMENT IN THE PRESENCE OF HUNTINGTON'S DISEASE

ÁLVARO GONZÁLEZ-REDONDO, FRANCISCO NAVEROS, EDUARDO ROS, JESÚS A. GARRIDO*
Department of Computer Architecture and Technology
University of Granada, Granada, Spain
https://doi.org/10.1142/s0129065720500574

The basal ganglia (BG) represent a critical center of the nervous system for sensorial discrimination. Although it is known that Huntington's disease (HD) affects this brain area, it still remains unclear how HD patients achieve paradoxical improvement in sensorial discrimination tasks. This article presents a computational model of the BG including the main nuclei and the typical firing properties of their neurons. The BG model has been embedded within an auditory signal detection task. We have emulated the effect that the altered levels of dopamine and the degree of HD affectation have in information processing at different layers of the BG, and how these aspects shape transient and steady states differently throughout the selection task. By extracting the independent components of the BG activity at different populations it is evidenced that early and medium stages of HD affectation may enhance transient activity in the striatum and the substantia nigra pars reticulata. These results represent a possible explanation for the paradoxical improvement that HD patients present in discrimination task performance. Thus, this paper provides a novel understanding on how the fast dynamics of the BG network at different layers interact and enable transient states to emerge throughout the successive neuron populations.

Keywords: basal ganglia; spiking neural networks; computational model; Huntington's disease; dopamine.

1. Introduction

Choosing the right action among many available options represents a primary but also challenging behavior for animal species. The BG have long been thought to play a pivotal role in the action selection process in the mammal brain.[1] A well-accepted hypothesis is that these nuclei choose between multiple motor commands coming from cortical areas.[2,3] However, it is still unclear how the BG filter incoming cortical commands in order to produce an accurate and fast output. This article aims to explore the signal processing in the BG by embedding a computational model of this brain area in a behaviorally relevant experimental setting involving action selection.

The BG network presents complex anatomical and functional sub-divisions, but it is usually structured in five main neuron populations[4] which can be organized into three sections (Fig. 1):

- The inputs of the BG are mainly received through the corpus striatum, with its main cell type being the medium spiny neurons (MSN), and the subthalamic nucleus (STN) neurons.
- The intermediate layers are composed by the external segment of the globus pallidus (GPe) and the substantia nigra pars compacta (SNc).

*Research Centre for Information and Communications Technologies (CITIC-UGR). Calle Periodista Rafael Gómez Montero 2, E18071 Granada, Spain. E-mail: jesusgarrido@ugr.es

- The output projection to the thalamus is finally carried by the substantia nigra pars reticulate (SNr).

The connectivity of these populations is mainly drawn according to three main routes from the cortex to the thalamus as follows (Fig. 1):

- The *direct pathway*, where the cerebral cortex makes excitatory glutamatergic synapses into the MSN_{D1}, which inhibits the SNr.
- The *indirect pathway*, where the cerebral cortex excites the MSN_{D2}, which inhibit the GPe, and finally, the GPe which also inhibits the SNr.
- The *hyper direct pathway*, where the cortex makes glutamatergic connections into the STN, which diffusely excites the SNr.

In addition to these broad pathways, there are dopaminergic projections from the SNc to the MSN, the STN and the GPe with modulatory effects (shaded box in Fig. 1).[1,5] Finally, the GPe forms recurrent loops with the STN.

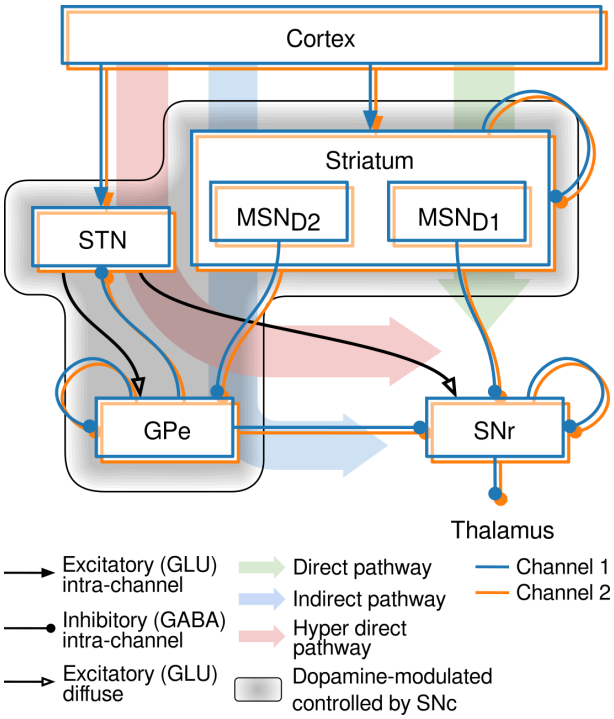

Figure 1. *Computational model of the basal ganglia.* Basal ganglia representation structured in channels (blue and orange) showing the direct, indirect and hyper direct pathways.

It has been hypothesized that the BG process a large number of cognitive streams or *channels* in parallel,[6,7] each of them representing a feasible action to be performed.[8] The BG is thought to act as an action selection machinery by inhibiting every non-selected action in the thalamus with the SNr, based on their corresponding activity level or salience.[3] A possible explanation for this mechanism was suggested by Ref. 1. They identified two steady-state and transient selection components, generated both in the striatum due to the cortex activity. According to this theory, the transient component in the striatum temporarily enhances the difference between several competing cortical inputs.

In order to shed some light on this action selection process, previous research in the literature has addressed both the natural and artificial alterations of the BG circuitry. For instance, the use of levodopa, a dopamine (DA) neurotransmitter precursor, modifies the levels of DA in the BG and systematically produces reduced reaction times and increased accuracy in simple auditory discrimination tasks in healthy subjects.[9] Additionally, several diseases can naturally affect the normal operation of the BG. This is the case of HD, which produces an enhanced activation of the N-methyl-D-aspartate (NMDA) glutamatergic receptors of the MSN in the striatum,[10] culminating in excitotoxicity (i.e. cell death).[11] Moreover, it is known that MSN expressing D2 dopaminergic receptors are more affected than those expressing D1 dopaminergic receptors in the early stages of HD,[12] disrupting the indirect pathway (MSN_{D2}-GPe-SNr). Finally, levodopa can potentiate the HD symptoms by exacerbating choreiform movements.[13] Despite these mainly negative effects of HD, some researchers have found a paradoxical improvement in auditory decision tasks (in both reaction time and accuracy) during early stages of the disease, presumably caused by the enhanced efficacy of NMDA receptors.[14] This improvement can help us to better understand both HD and the action selection process in the BG.[1]

Although different non-invasive experimental techniques and statistical analyses, such as electroencephalogram and information metrics, allow the identification of important brain areas related to specific tasks,[15,16] the way in which they contribute to complex behaviours remains highly elusive. In recent decades, the use of biologically inspired computational models emulating spiking neu-

ral networks[17] has been demonstrated as being useful for understanding experimental recordings from multiple brain areas[18,19] and for studying different neurological alterations.[20–22] Thus, computational models represent a promising approach to explore not only the normal operation of the BG, but also how different artificial alterations (e.g. levodopa) or diseases (e.g. HD[1] or Parkinson's disease (PD)[23,24]) can affect this nucleus. Although many computational models of the BG have been proposed to explain the overall operation of this brain area,[6,25–29] it remains unclear how the transient phenomena generated by the MSN[1] in the striatum on HD patients propagates to the SNr (BG output nucleus), how this transient phenomena facilitates action selection, and how DA affects this process.

In this paper we present a computational model of the BG integrating all its main neuron types. This model facilitates the exploration of the emergence of both steady and transient phenomena in the MSN and the interplay between the three BG pathways in the propagation of these phenomena. In this framework we have been able to quantify the selectivity between competing actions transmitted to the thalamus from the SNr output. In addition to this, we have explored how altered conditions, such as increased DA levels or the alterations produced by HD, affect the performance of the BG as selection machinery using a stimulus discrimination task as a test-bench.

Section 2 provides details on the implementation of the computational model. Section 3 describes the results emerging from the simulation of the computational model in the framework of the stimulus discrimination task. In section 4 we discuss these results regarding previous computational models and experimental evidence in the literature. Finally section 5 summarizes the main contributions of this article.

2. Computational modeling of BG and HD

A computational model of the BG, including all its main nuclei and connections, has been implemented (Fig. 1). The network structure and the neuron models used in this manuscript are based on recent work by Fountas and Shanahan.[30] The model includes five neuronal populations and nine neuronal types (all of them implemented as Izhikevich neuron models, but with different parameters in order to capture their particular cell dynamics). The total number of simulated neurons is 5,494 divided as follows: the MSN layer contains 2,292 neurons, with half of them (1,146) expressing D1 receptor and the other half D2 receptor. The STN layer contains 47 neurons, the GPe 155 neurons, and the SNr 3,000 neurons.

The neuron populations in our BG model have been connected following a channel structure. As a general norm, the neurons in every channel are only allowed to synapse neurons in the same channel. The exceptions are the STN efferents, which are diffuse and connect to all the channels (Fig. 1), and the lateral inhibition within and between the MSN channels and SNr channels. The modulatory connections from the SNc are considered implicitly as the global level of tonic DA in the model. The average level of activation (i.e. the firing rate) in each channel at the MSN represents the salience or urgency of the action represented at that particular channel.[31]

For the proposed selection task, we have implemented three different channels in our BG model following Ref. 1: one for the selected option (with 40% of the neurons), one for the non-selected option (with 40% of the neurons), and the third not competing in the selection task (with 20% of the neurons). The third channel represents background neuronal activity able to influence the other channels through the diffuse connectivity from STN and the lateral competition in MSN and SNr. For the sake of simplicity, only the selected (blue) and non-selected (orange) channels have been represented in the figures.

The following sub-sections describe the behavior of the neuron models used, and the appendices go deeper into the modeling details. The source code of the model implementation for NEST 2.12,[32] as well as the scripts allowing the reproduction of the results shown in this article, have been made available at the following address https://github.com/EduardoRosLab/BG_selectivity.

2.1. *Neuron and synapse models*

The Izhikevich neuron model[33] has been chosen to reproduce the experimental firing modes recorded in the different neuron types of the BG. The parameters for each neuron type have been optimized following the adjustment procedure described in Ref. 34. This method aims to approximate electrophysiological properties (e.g. the action potential amplitude and width, the resting and threshold potentials and the rheobase current) and their

steady-state frequency-current (F-I) relations. Figure 2 shows the reference and resulting F-I curves for each neuron type. We selected the parameters of our MSN (see https://github.com/EduardoRosLab/BG_selectivity/raw/master/parameters_tables.pdf in the repository) from different sources in order to obtain a good match between experimental data and simulated behavior, namely transient selectivity (see below). When the rest of the BG nuclei were added, their neuron model parameters were calculated following the previously described parameter estimation procedures[34] or by local search/manual tuning.

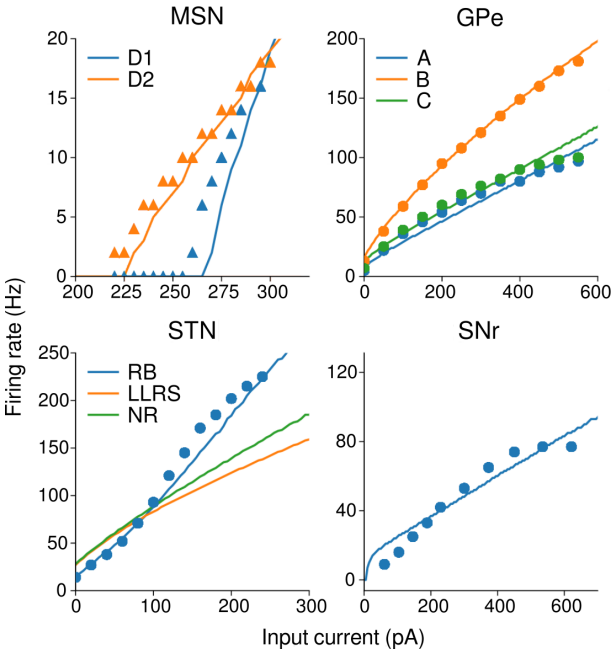

Figure 2. *Frequency-current (F-I) curves.* Solid lines represent our computational results, while triangles and dots respectively represent the simulated and experimental data used to tune our models.

More than 90% of the striatal neurons are MSN, showing competitive behaviors between channels through lateral inhibition, both directly and through interneurons.[35] We modeled the striatum as a population of MSN with lateral inhibition. These neurons show characteristic firing patterns such as long-latency first spike following current injection or membrane potential bi-stability in response to random input activity, with a hyperpolarized *down-state* and depolarized *up-state* plateau.[28] Although other neuron types have been reported (e.g. diverse GABAergic populations[36] and cholinergic interneurons with a role in reinforcement-related signals[37]), we have intentionally ruled them out since our preliminary simulations showed that they did not impact the transitory effect under study in the proposed experimental setup. The MSN are divided into two sub-populations of 1,146 neurons expressing different types of DA receptors (D1 and D2). Thus, two neuron types have been adjusted for the MSN subpopulation (see appendix 1 for further details). Figure 2 shows the comparison of the F-I curves from our Izhikevich neuron models and the highly-detailed multi-compartment models of the MSN.[27]

Neurons in the GPe have shown at least two different firing patterns in primates: high-frequency discharge separated by intervals of total silence (HFD) and low-frequency discharge and bursts (LFD).[38] Interestingly, similar intracellular recording in rats[39] have been reported to show three different identifiable firing patterns. In our model, we have followed this latter approach by including three neuron types named A, B and C.[34] Our model includes 131 neurons of type B, which behave similarly to HFD neurons (the only neuron type able to evoke rebound firing), and 7 and 17 neurons of types A and C respectively, which behave similarly to LFD. Figure 2 shows the matching of the simulated neurons and the experimental data (dotted) from Ref. 39.

The STN is composed of three different neuron sub-types. All of them behave similarly when depolarized, with sigmoid F-I relation.[40] However, they have shown different responses after long depolarization, including rebound bursts (RB), long-lasting rebound spikes (LLRS) and no rebound effect (NR). Our model respectively includes 28, 12 and 7 neurons of each cell type distributed between the three channels. Figure 2 includes experimental data (dots) for STN RB cell type from Ref. 40.

Finally, the GABAergic SNr neurons show spontaneous high-frequency firing that may turn abruptly into bursting or silence depending on external input. In addition to this, this type of neuron emits rebound spikes.[41,42] Our model includes 3,000 SNr neurons, whose parameters have been adjusted to obtain firing rates around 20Hz in absence of external current stimulation, and silent when the channel is selected. These firing rates fall within the range obtained in cell recordings considered in other computational models.[37,43,44] Figure 2 shows a comparison

of the firing rate of the simulated neuron and the experimental data for the SNr neuron type from.[45]

Table 1. Synaptic and connectivity parameters.

Connection	Receptor	Connectivity	Probability
Cortex → MSN	AMPA	Intra-channel	1.0
	NMDA	Intra-channel	1.0
MSN → MSN	$GABA_A$	Inter-channels	0.32
MSN → SNr	$GABA_A$	Intra-channel	0.033
SNr → SNr	$GABA_A$	Inter-channels	0.1
Cortex → STN	AMPA	Intra-channel	1.0
	NMDA	Intra-channel	1.0
STN → GPe	AMPA	Inter-channels	0.3
	NMDA	Inter-channels	0.3
GPe → STN	$GABA_A$	Intra-channel	0.1
GPe → GPe	$GABA_A$	Intra-channel	0.1
MSN → GPe	$GABA_A$	Intra-channel	0.033
STN → SNr	AMPA	Inter-channels	0.3
	NMDA	Inter-channels	0.3
GPe → SNr	$GABA_A$	Intra-channel	0.1066

All the neurons included in our BG model implemented chemical synapses. These were modeled using synaptic conductances governed by exponentially-decaying functions and constant synaptic weights (without plasticity mechanisms). Three types of chemical receptors with different temporal dynamics were implemented: AMPA, NMDA and $GABA_A$. The NMDA receptor also models the voltage-dependent magnesium plug.[46] In Table 1 we show the interconnectivity topology of our BG models. All the neurons implemented a probabilistic all-to-all connectivity distribution with connectivity ratios between neuron types extracted from literature.[1,34] These connections could be intra-channel (neurons just connected with neurons in the same channel) or inter-channels (neurons connected with neurons in the same or different channels). The rest of synaptic parameters were selected from the literature or obtained from local search/manual tuning. Details on the implementation can be found in appendix B.

2.2. *Huntington's disease modeling*

HD has been demonstrated to disrupt the indirect pathway of the BG by reducing the number of MSN D2 neurons[12] during the early stages of the evolution of the disease. In our study we have modeled this effect by randomly removing a fraction of the MSN D2 neurons (Eq. (1)). Additionally, MSN D2 neurons also over-express NMDA receptors in early symptomatic and pre-symptomatic HD patients, leading to excessive action-potential emission and eventually neuronal apoptosis.[47] This effect has been modeled by increasing the synaptic weight of NMDA receptors onto the MSN neurons (Eq. (2)).

$$n_{MSN-D2} \leftarrow n_{MSN-D2}(1 - hd \cdot a_r) \quad (1)$$

$$w_{NMDA} \leftarrow w_{NMDA}(1 + hd \cdot s_r) \quad (2)$$

where hd represents the level of HD ranging from zero (no HD effect) to one (maximum considered HD affectation), n_{MSN-D2} is the number of MSN neurons with DA receptor D2, a_r is the cell apoptosis ratio ranging from zero (no apoptosis) to one (maximum apoptosis), w_{NMDA} is the synaptic weight from cortex to MSN, and s_r is the NMDA increase factor. The a_r and s_r parameters have been set to 0.8 and 1.0 since these values maximize the network effect during selection tasks according to previous research.[1] The combination of both effects simulates the early and middle stages of HD (grades one and two in the neuropathological scale proposed by Ref. 12). Finally, since advanced stages of HD are incompatible with behavioral experimentation (HD patients lose their motor control capabilities), modeling further damages in our BG model due to advanced stages of HD[48] remains beyond the scope of this work.

3. Results

3.1. *Experimental framework*

HD patients show a paradoxical improvement (both in speed and precision) in the auditory decision task proposed by Beste.[14] In this experiment the subject is told to distinguish between short (200ms) and long (400ms) auditory tones in a series by pressing a left or right button for each option (Fig. 3A).

The main inputs to the MSN and STN come from afferent axons of the pyramidal neurons in layer V from different cortex areas[49] (e.g., the auditory cortex[50]). This model assumes that the decision on the presented tone length, and its consequent motor action, has previously been performed in the cerebral cortex and propagated to the BG by modifying the firing rate of the input fibers that arrive to each BG channel. This cortical input activity has been emulated by means of three populations of Poisson spike train generators emulating the cortical activity. These three populations project over the three channels in the MSN and STN (two channels for the left and right motor response, and a third chan-

nel processing cortical activity not related with the task). Thus, each neuron in MSN and STN receives the equivalent of 250 randomly-chosen input spike trains.[29]

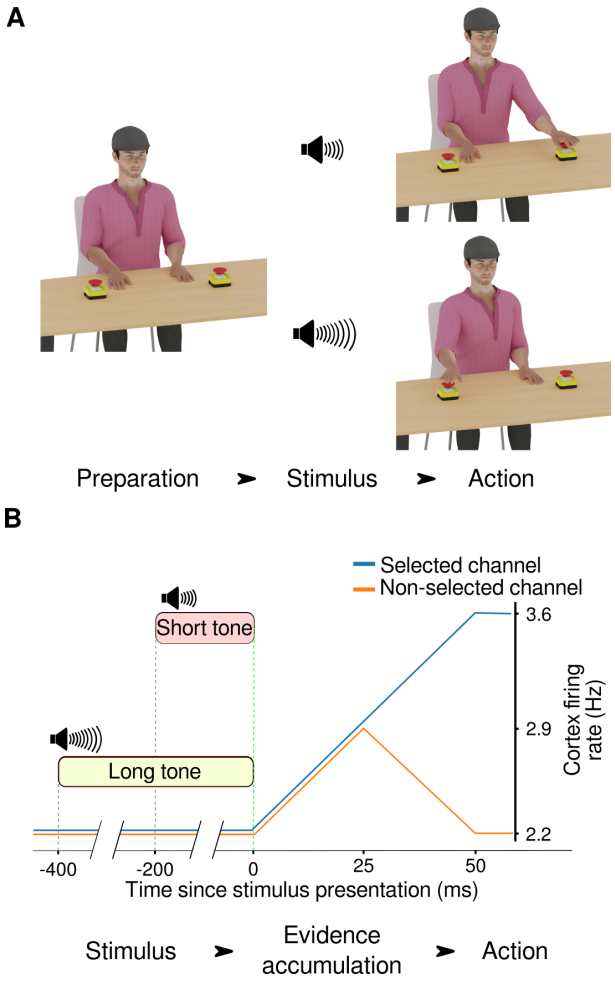

Figure 3. *Experimental framework.* **A.** Stimulus discrimination experimental procedure: the subject must press the left (right) button straight after a short (long) tone onset. **B.** Firing rate evolution in the selected and non-selected channels in the cortex.

The simulated protocol was taken from Ref. 1. It started with a stabilization period of 1,500ms in which the mean firing rate of the three Poisson populations was fixed to a baseline activity of 2.2Hz. The end of this stabilization period corresponded with the end of the auditory stimulus. During the following 25ms, the two Poisson populations corresponding with the short and long tones (press left and right button) gradually increased their mean firing rate to a medium excitation level of 2.9Hz (Fig. 3B). After this 25-ms period, the subject was able to discriminate the tone length and select the corresponding motor action. During the next 25ms, the Poisson population corresponding with the selected action kept increasing its firing rate until it reached an average firing-rate of 3.6Hz. On the contrary, the non-selected population returned to the baseline level of activity. These activation levels were maintained during an additional period of 1,000ms. Finally, the third Poisson population remained in the baseline state during the whole experiment (2,550ms). All these firing rates range within reported biological constraints for the auditory cortical layer.[51] Note that with this experimental protocol we were not modeling the detection of longer tones (which were already encoded in the incoming cortex activity), but rather the selection of the action to be performed in response to the detection of short or long tones.

3.2. *Data analysis: selectivity metrics*

The resulting action potentials obtained during the simulation of the model (each experiment lasting 2,550ms) were used to generate the activity histogram (1-ms bin) for each channel in every neuron population. The population spikes were filtered by convolving with a 7.5-ms Gaussian kernel to mimic the resulting excitation/inhibition received in the successive layer. Aiming to rule out the high variability of the resulting activity histograms, the instantaneous firing rate in each time bin has been averaged over 40 simulations with different random seeds for each experimental condition (Fig. 4C).

In order to provide a quantitative evaluation of the performance of the computational model in the proposed behavioral task, the following assumption has been made, which has been widely hypothesized before:[2,6,7,52–54] the BG chooses the action with the highest reward expectance between multiple possible actions by increasing (decreasing) the activity in the corresponding MSN (SNr) channel and reducing (maintaining) the firing rate in the remaining MSN (SNr) channels. Thus, the following estimators aim to quantify how distinguishable the activity profiles of the MSN and the SNr are in the considered channels.

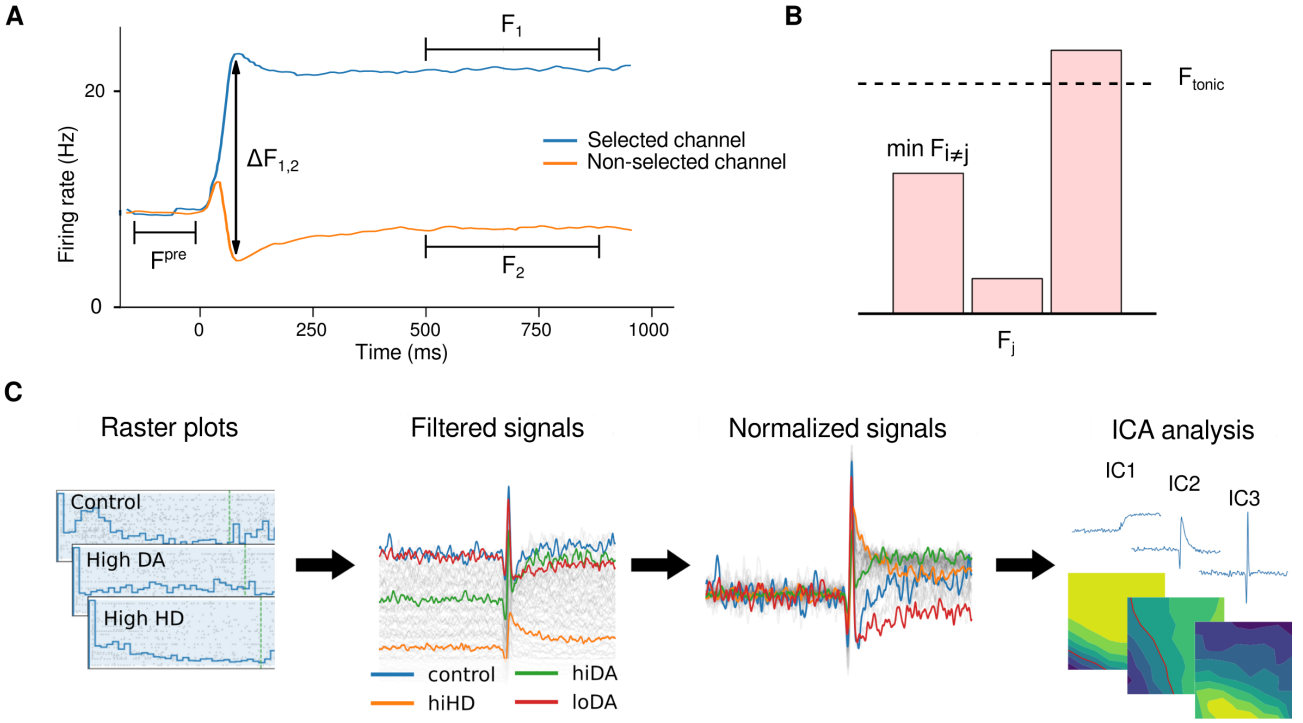

Figure 4. *Analysis performed on the simulations results.* **A.** Selected **(blue)** and non-selected **(orange)** channel activities after the application of Gaussian kernels to the raster plot histograms (MSN or SNr). **B.** Mean activity levels of the SNr channels before the tone onset (F_{tonic}) and after the tone onset (F_i, F_j), where F_j is the selected channel and F_i the non-selected one. **C.** For the ICA analysis, the raster plot histograms are filtered, normalized and decomposed in their ICA components with their corresponding weights.

Activity evaluation in the striatum: selectivity

Following the approach in Ref. 1, the *selectivity* in the MSN can be defined as the ability to robustly distinguish competing signals. Two complementary modes of selectivity have been proposed, measured with different metrics applied to the mean activation of each MSN channel population. Given a competition between different cortical inputs, a transient selectivity is the temporary promotion of the most salient signal simultaneously to suppression of the least-salient signal. This effect results in the transient boost of the difference in salience between the competing signals. The transient selectivity in the MSN (TS_{MSN}) is defined according to

$$TS_{MSN} = 1 - \frac{\overline{F_1} - \overline{F_2}}{\Delta F_{1,2}} \tag{3}$$

where F_1 and F_2 are two signals (firing rates of the two competing channels), $\Delta F_{1,2}$ is the maximum difference between F_1 and F_2 occurred during a transition window between 0ms and 200ms after the stimulation, and $\overline{F_1}, \overline{F_2}$ are the mean stable activity of signals F_1 and F_2 after the transition period (Fig. 4A).

Moreover, given a competition between different input signals, the least salient signal (in our experiments, F_2) tends to be inhibited on a sustained basis by the most salient signal (Fig. 4A); this is called the *stationary selectivity* (SS_{MSN}) and is defined as follows:

$$SS_{MSN} = 100 \times \left(1 - \frac{\overline{F_2}}{\overline{F^{Pre}}}\right) \tag{4}$$

where $\overline{F^{Pre}}$ is the mean stable activity of the signal F_1 or F_2 before the tone onset (both activity level are similar before the stimulus onset). Thus, the TS_{MSN} provides an estimation of how distinguishable the competing signals are during the transitory state while the SS_{MSN} quantifies how distinguishable the competing signals are once they have reached their steady state.

Activity evaluation in the SNr: distinctiveness

A more general metric, which can be applied to the SNr, is the distinctiveness of a single selected channel, defined as the ability of a channel to generate distinctively less activity than any other channel in the layer.[34] Since the SNr inhibits the thalamus, the distinctive channel (the selected one with the lower activity) inhibits the corresponding channel in the thalamus in a lesser degree, propagating the BG selection to the thalamus. The distinctiveness in the SNr represents the degree to which the following two conditions are fulfilled: (a) the firing rate of the selected channel in the SNr is close to zero, and (b) no other channel is far below tonic levels. These two conditions can be quantitatively evaluated over the time t according to $a_j(t)$ and $b_j(t)$ respectively:

$$a_j(t) = 1 - \frac{F_j(t)}{\max\{F_{tonic}, F_j(t)\}} \tag{5}$$

$$b_j(t) = \frac{\min F_{i\neq j}(t)}{\max\{F_{tonic}, \min F_{i\neq j}(t)\}} \tag{6}$$

where j is the examined channel, $F_j(t)$ is the firing rate of channel j at time t, $\min F_{i\neq j}(t)$ is the minimum SNr firing rate of any channel different to j at time t and F_{tonic} is the tonic firing rate of the SNr, assumed here to be 20spikes/sec (Fig. 4B). Then, the distinctiveness $D_j(t)$ is defined as:

$$D_j(t) = a_j(t) \cdot b_j(t) \tag{7}$$

with $D_j(t)$ values range in $[0, 1]$, with 1 indicating that the channel j at time t propagates distinctively less inhibition than any other channel to the thalamus, and 0 the opposite condition (or channel activity j is far from zero i.e. it is not chosen, or some other channel is closer to zero i.e. the other channel is chosen instead). The *steady-state distinctiveness* and *transient distinctiveness*[55] are both calculated from $D_j(t)$. The former is calculated as the average of the stable post-transient activity (the signal is assumed to reach steady state after 500ms from the stimulus onset) while the latter is defined as the maximum distance between the distinctiveness of the channels during a fixed short interval (200ms in our experiments) after the generation of the salient signal.

Independent component analysis

In order to evaluate the different temporal components emerging from the population activity, we have applied the independent component analysis (ICA) algorithm[56] to the filtered signal in the selected channel of the MSN (Fig. 4C). ICA is a widely used computational method for separating a multivariate signal into its additive non-orthogonal components. This is done by assuming that the subcomponents are non-Gaussian and statistically independent signals. Although this algorithm is similar to other classic methods, such as principal component analysis (PCA), ICA imposes to the resulting signals the harder constraint of being statistically independent (and not just linearly uncorrelated as in PCA). Based on preliminary simulations within the experimental setup under evaluation in this article, ICA demonstrated to be more successful than the PCA on finding significant components.

Prior to the application of the ICA algorithm, each signal was normalized by subtracting the mean activation level before stimulus onset and the resulting signal was divided by its standard deviation, so that it becomes insensitive to any possible firing rate additive or multiplicative variation. After that, we used the *FastICA* decomposition function from the SciKit-Learn Python library[57] to obtain the independent components from the signals. We chose the number of independent components to be at least two (as we make the assumption of the existence of the transient and the steady-state components) and as high as needed to be able to explain at least 90% of the variability of the signal. The same analysis was applied to both the selected and the non-selected channel populations of the SNr. Finally, the ICA algorithm provided the relative weights of each component in each experimental condition. The resulting weights allowed us to evaluate the presence and the relative importance of each temporal component for different combinations of factors.

3.3. *General network behavior*

Our network model was first simulated in control conditions (no HD $hd = 0$ and default level of DA $d_1 = d_2 = 0.3$) during Beste's task.[14] The resulting activity of each population and their respective channels are shown in Figure 5. This activity falls within *in vivo* values in all nuclei: the MSN firing

rates are below 2Hz when low activity and above 17Hz in high activity.[34,44] The STN normally fires at around 10Hz, but can get as high as 30-50Hz.[58] The GPe firing rate is around 30Hz without activation but raise to 40Hz when its channel is not selected and decrease to almost zero when its channel is selected.[34] The SNr activity has been reported to be close to zero when receiving inhibition from MSN D1 and around 20-30Hz when it is being activated.[34,44] Not surprisingly, the variability (standard deviation) of the population firing rate depends on the number of neurons included in each nucleus (ranging from 46 neurons in the GPe to 3,000 neurons in the SNr). The initial 1,500ms are devoted to stabilizing the network activity in response to the basal activation in the cortex (Figs. 3B and 5). Some nuclei, such as the GPe and the SNr, show intrinsic activation, resulting in high firing rates at the beginning of the simulation that slowly decrease due to the lateral inhibition existing within each nucleus (Fig. 5). On the contrary, the MSN demonstrated the slowest adaptation mainly due to the intrinsic long first-spike latency of this neuron model.

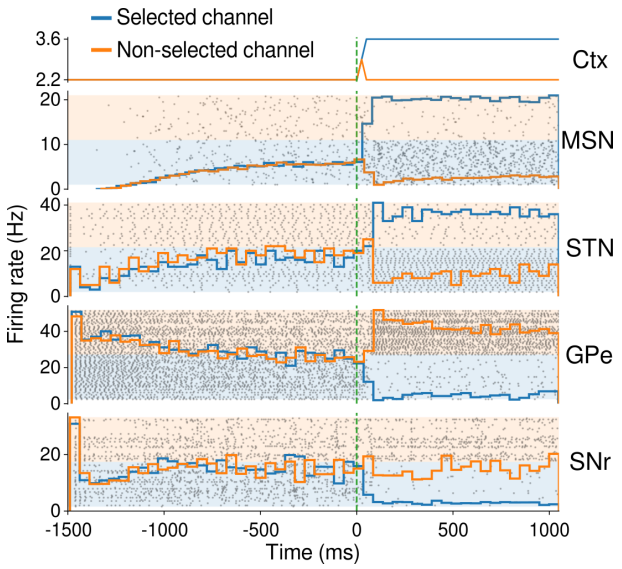

Figure 5. *General network behavior.* Raster plot (black dots) and population firing rates (solid lines) for the cortex (Ctx), MSN, STN, GPe and SNr populations. Selected and non-selected channels are respectively drawn with blue and orange background/line colors.

Once the network activity becomes steady, the cortex increases its activity in both channels (cortex selection onset) (Fig. 3B). After 25ms, the cortex selects only one channel (which further increases the firing rate, while the other one returns to the basal level of firing) (Fig. 3B). This channel selection in the cortex produces a strong response in the BG: the selected channel increases their firing rate in the MSN and STN and inhibits, in turn, the non-selected channels (Fig. 5). However, the selected channel in the MSN shows transient phenomena by producing peaks of activity (200ms long or less, as will be discussed later in Fig. 6B) due to the intrinsic properties of its neuron model, the long time constant of the NMDA receptor and the lateral inhibition. Conversely, the selected channel in the GPe and the SNr receives strong inhibition from the MSN, thus compensating for the stronger excitation from the STN and leading to a notorious reduction of the firing rate in the selected channel of the GPe and SNr (Fig. 5). Due to the recurrent loops between the GPe and STN (Fig. 1), the activity of the GPe remains unstable for 500ms after the cortex selection onset. The output nucleus (SNr), not unexpectedly, decreases the activity for the selected channel, while the non-selected one remains with a basal level of activity (after a transient activity peak) (Fig. 5).

The commented network operation is valid for the control case (default DA and no HD affectation). We also tested the network with different levels of DA or HD. Increased DA levels in the model resulted in enhanced response of the MSN to the cortical input, as previously reported in freely moving rats.[59] In addition to this, the firing rates obtained for control and pathological HD MSNs are in agreement with the experimental results obtained from mice.[60] We used our model to explore if altered levels of DA and the presence of HD affectation may change the balance between excitation and inhibition in any network layer and, as a consequence, if they produce enhanced/reduced levels of selectivity in the MSN and distinctiveness in the SNr.

3.4. *Striatum (MSN) activity*

Overall, our simulations demonstrated that increase in either DA levels or noteworthy HD affectation (or both conditions together) resulted in enhanced levels of MSN activation (Fig. 6A green, orange and purple solid lines in the top plot) and the rest of the BG nuclei, resulting in different balances of excitation/inhibition depending on the particular configuration.

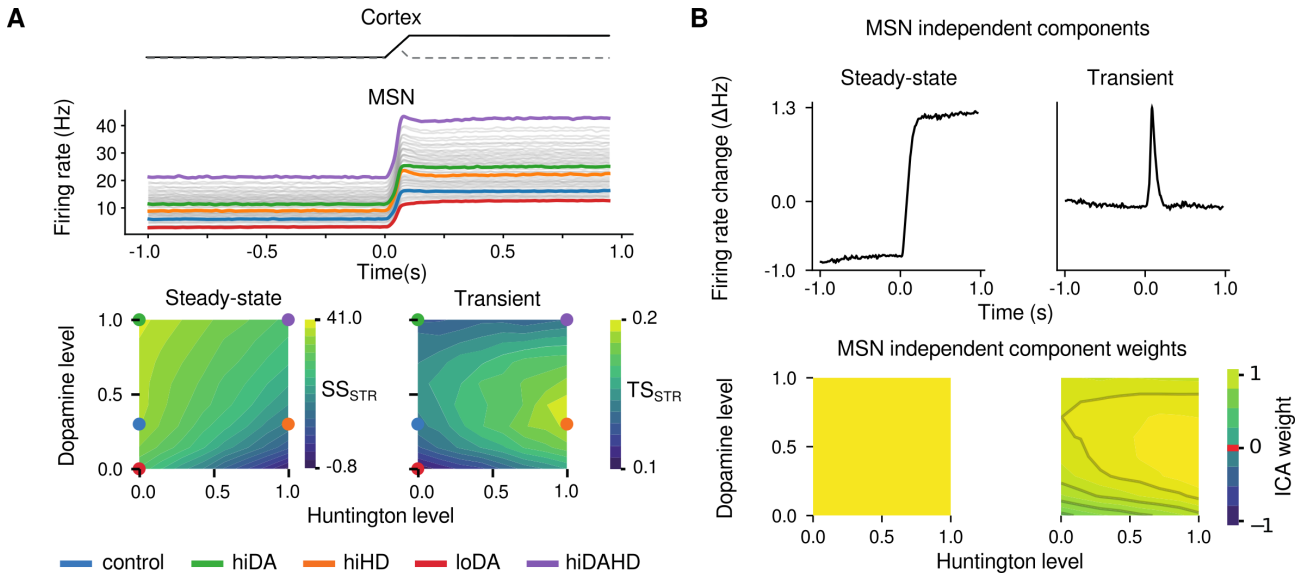

Figure 6. *Effect of different dopamine (DA) levels and Huntington's disease (HD) conditions in the medium spiny neurons (MSN).* **A. (Top)** Average firing rate of the selected channel in the MSN. Each trace represents a different setting, while some representative conditions have been highlighted in colors: *control* (default DA level and no HD) (blue), *hiHD* (default DA and high HD) (orange), *hiDA* (high DA and no HD) (green), *loDA* (low DA and no HD) (red) and *hiDAHD* (high DA and HD) (purple). **(Bottom)** Steady-state (SS_{MSN}) (left) and transient (TS_{MSN}) (right) selectivity within the studied parameter space (DA vs HD level). Colored circles mark the cases previously considered. **B. (Top)** Independent components obtained from the MSN firing histogram by using the ICA algorithm over all the experimental conditions. **(Bottom)** Weight of each signal component in each experimental condition as obtained from the ICA algorithm.

In order to achieve a fuller understanding of the functional effect that altered DA and HD levels produce in the processing layers of the BG, we simulated a whole set of different configurations of the network to perform Beste's task. For each experimental condition, the activity histogram of the selected channel in the MSN has been extracted (top plot in Fig. 6A) and the steady-state and the transient-state selectivity have been analyzed (bottom plots in Fig. 6A). As a general rule, steady-state selectivity is enhanced by increased DA levels while HD affectation reduces it. On the other hand, transient selectivity is increased by HD affectation and shows an inverted "U" relationship with DA levels (medium DA levels resulted in increased transient selectivity).

Aiming to discriminate the effect of DA and HD in the emergence of steady or transient components of activity, we applied the ICA algorithm on the activity histograms of the selected channel in the MSN, resulting in the two components indicated in the top row plots in Figure 6B. The first component (left plot) represents a steady signal (corresponding with the steady state before and after stimulus onset), while the second component (right plot) shows transient behavior around the stimulus onset. By exploring the weights associated to each component for each experimental condition, the first component is similarly present in all the experimental conditions, while the second component shows differences among the configurations (bottom row in Fig. 6B). The second component is more prominent with high HD affectation and medium DA levels.

3.5. *Substantia nigra (SNr) activity*

Similarly as previously analyzed in the MSN, we have explored the averaged firing activity of the SNr (Fig. 7A) in response to the cortex input activity that mimics the stimulation received during Beste's task. Moreover, we have also analyzed the distinctiveness[34] of the stimulus with different network conditions (DA and HD levels). The results show that both the transient and the steady-state distinctiveness increase with high levels of DA and reduced HD affectation.

We also used ICA in the SNr signals, obtaining three independent components able to explain at least the 90% of the variability of the original signal. Fig. 7B shows the extracted independent com-

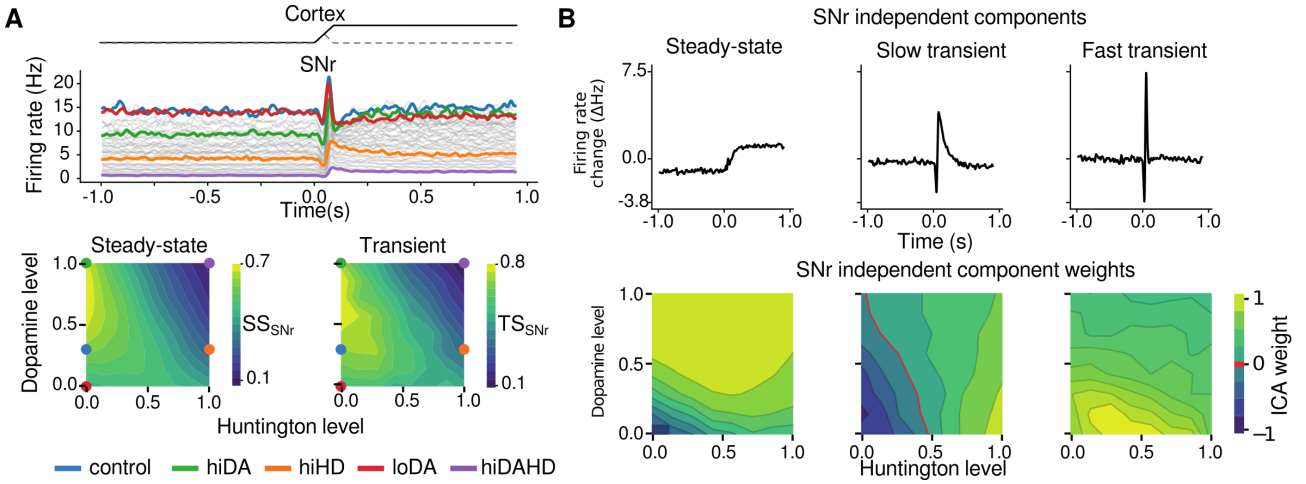

Figure 7. *Effect of different dopamine (DA) levels and Huntington's disease (HD) conditions in the substantia nigra reticulata (SNr).* **A. (Top)** Average firing rate of the non-selected channel in the SNr. Each trace represents a different setting, while some representative conditions have been highlighted in colors: *control* (default DA level and no HD) (blue), *hiHD* (default DA level and high HD) (orange), *hiDA* (high DA level and no HD) (green), *loDA* (low DA level and no HD) (red) and *hiDAHD* (high DA level and HD) (purple). **(Bottom)** Steady-state (left) and transient (right) distinctiveness in the SNr within the parameters space studied (DA vs HD level). Colored circles mark the cases previously considered. **B. (Top)** Independent components obtained from the SNr firing histogram by using the ICA algorithm over all the experimental conditions. **(Bottom)** Weight of each signal component in each experimental condition as obtained from the ICA algorithm.

ponents and their corresponding weights for each experimental condition. The first component is associated to the steady-state evolution, while the other two are related to the transient phenomena (slow and fast). The distribution of the weights of these components shows very distinct patterns. While the steady-state and slow transient components show some linearity (the first with DA level and the second with HD level), the fast transient component shows a non-linear behavior, with more weight when the DA level is low and the HD level is medium.

4. Discussion

4.1. *Interpretation of results*

The obtained results are in agreement with previous research. Ref. 1 showed that the potentiation of the transient component explains (at least, partially) why HD patients achieve better performance in timed decision tasks. This effect was previously explained as enhanced information processing in simple sensorimotor tasks. By using ICA we have evidenced neuronal correlates of this experimental performance improvement resulting from increased DA levels in healthy subjects receiving levodopa.[9] Specifically, the transient independent component detected at MSN is more prominent with high HD affectation and medium DA levels.

Regarding the SNr activity, both the steady-state and the transient distinctiveness fail to explain the enhanced performance of HD patients in Beste's task. Although previous articles in the literature have associated this paradoxical improvement to the alteration of the selectivity in the MSN, our simulations demonstrate that increased selectivity does not propagate to the subsequent SNr layer (the output nucleus of the BG). By studying the independent components obtained from the activity histograms we are able to offer an alternative explanation to this paradoxical improvement. After observing the shape of each component (bottom plots in Fig. 7B) we have concluded that these components can propagate from the MSN to the SNr. The weight of each component for each experimental condition indicates which circumstances facilitate the propagation of the corresponding component (e.g. while the slow-transient component reliably propagates with high HD, the fast-transient component more consistently propagates with medium HD and low-medium DA). Thus, the propagation of the fast-transient component from the MSN to the SNr may support the paradoxical improvement in Beste's task observed in patients in the

early stages of HD.

4.2. *Comparison with previous studies*

To date, several BG computational models have been proposed for different purposes. A detailed[27] and a subsequently simplified[28] computational model expressing the DA modulation in MSN D1 and D2 were proposed, and then added to a three-dimensional network model together with fast spiking interneurons (FSI) in the striatum.[29] Other models studied the exploration/exploitation trade-off,[8,30,61,62] or reproduced diverse behavioral tasks with a complex model containing several neuronal nuclei (cortex, BG and thalamus).[63] In some models, phasic DA signals have been added on top of a tonic DA value, reproducing the neural mechanism for which the triggering of a movement requires a dopaminergic burst just preceding the movement onset.[64,65] There are studies, more focused on HD, as in Ref. (1), where they studied the origin of HD paradoxical effects as a consequence of the alteration of the transient selectivity in the MSN.

The general behavior of the MSN population has been evaluated against selectivity metrics previously proposed in the literature.[1] According to these metrics, high levels of HD affectation present contradictory effects (decrease and increase) in steady-state and transient-state selectivity, in agreement with previous simulations. These results confirm the prediction that the transient-state selectivity metric in the MSN may explain the paradoxical speed improvement in Beste's task by HD patients,[14] assuming that middle levels of HD in our model correspond to the early stages of HD patients.[10,66] Although our simulations support this prediction, the evaluation of similar metrics in the subsequent layer (the SNr) indicates that the transient distinctiveness does not reflect the paradoxical behavioral improvement in HD patients. However, the analysis of the components extracted by the ICA algorithm in the SNr activity (the BG output layer) evidenced two transient components and one steady component. The weights of the components in HD conditions indicate that only the fastest component supports the paradoxical speed improvement in Beste's task. This component would act by abruptly avoiding the activity of alternative behavioral options.

Our computational model also allowed the analysis of tonic DA effect on BG operation. Previous studies did not specifically address the effect of DA in steady-state or transient-state selectivity (e.g. Ref. 1 included DA in the computational model of HD but with a fixed value throughout all experimental conditions). In our simulations, only the transient-state metric in the MSN evidenced decremented selectivity caused by high or low (non-medium) DA levels. According to these simulations, medium levels of DA may improve the subject's performance in selection tasks. Similar metrics in the SNr show enhanced steady-state and transient distinctiveness linked to higher levels of DA. These results are supported by the cognitive improvement registered in behavioral tasks by subjects receiving levodopa.[9] The ICA algorithm used on the SNr signals shows that transient components also occur in the SNr for high HD affectation (slow component) or a combination of medium HD affectation and low DA levels (fast component). These results explain how the augmented transient selectivity associated to the MSN of HD patients propagates to the SNr, projecting to the cortex through the thalamus and originating behavioral effects.

The application of the ICA algorithm in the SNr also evidences how different conditions affect each component of the signal. The steady-state component mainly depends on DA level, making this component a candidate for a non-pathological improvement mechanism in performance during selection tasks. It is in agreement with the experimental improvement of healthy subjects with high DA levels in selection tasks.[9] The slow-transient component is affected by the HD affectation but not by DA levels while the fast-transient component requires medium-low levels of DA and medium levels of HD affectation. Since paradoxical improvement on HD patients requires medium levels of HD affectation[14] and low or normal levels of DA, the fast-transient component closely fits this experimentally observed pattern. Thus, this fast component could be considered as a plausible marker for sensorial discrimination performance. Our model is also compatible with the reported deterioration of HD patients treated with levodopa,[13] where high HD and DA levels would deteriorate performance as the fast-transient component is reduced. In any case, further experimental studies are required in order to validate this hypothesis.

4.3. *Model limitations and future work*

One of the main limitations of the proposed model is that it lacks the recurrent loop between the BG and the cortex (where the decision process is thought to take place) through the thalamus. In absence of the cortico-BG-thalamic-cortical loop, our model is assuming a simple cause-effect relationship between the cortex and the BG. Current research considers a system-level approach where specific behaviors are generated by the interplay of different subsets of components of the brain.[67–69] The BG is not making the decision in isolation as the cortex is also taking part in this process: cortical feedback projections to the striatum and STN make the internal competition between channels a cumulative dynamical process.[70] Because of this, our result analysis is restricted to a small time window around the stimulus (before the re-entrant signal from the cortex is able to affect the BG activity). This time window is wide enough to allow us to explore the propagation of the transient components through a more extensive model of the BG than any previous research. Integrating the whole closed-loop is key in future research, as recently proposed.[64,65] The inclusion of cortical processing structures in a closed loop could facilitate further understanding of larger-timescale motor phenomena, such as the mechanisms of event-related desynchronization/synchronization found during motor imagery tasks.[71]

For the sake of simplicity of the analyses carried out in this study, we just took into account the tonic DA signals in our simulations (phasic DA signals were simplified). Future approaches should address how phasic DA signals might unbalance the state of equilibrium between the direct and indirect pathways.[64,65,70] Moreover, the presence of DA-dependent plasticity[26,31] in the cortex-MSN connections may somehow affect the BG processing of the incoming decisions. Nevertheless, the proposed model does not consider learning at any level (cortical or sub-cortical). Our simulations assumed the subject had previously learnt the action-selection task and it was on the automatization phase in which the cortex-striatum network plays a pivotal role.[72]

Although interneurons in the striatum (and specifically the FSI) shape the activity of the MSN,[36] and other models in the literature have already included these type of neurons,[1,34,35] we have avoided including these kinds of neurons as our preliminary simulations have shown no relevant effect for our particular behavioral task. This might happen due to the relatively low levels of input activity we have used in our experiments. The FSI show a stronger influence when a higher baseline and stepped inputs are used.[1]

Finally, one possible use of the components found in this study is to help us understand the origin of neurophysiology data obtained in real behavioral experiments. In Ref. 14 HD patients and controls were required to differentiate between short and long tones (a task very similar to the one simulated in this article) while an electroencephalogram (EEG) was recording the brain activity. They found that the paradoxical behavioral improvement (reflected as better accuracy and faster time responses in HD patients) correlated with the intensity of an event-related potential (ERP) signal obtained in the EEG known as mismatch negativity (MMN). This ERP signal could indicate the recognition of unexpected events by the auditory system.[14] Specifically, its presence can be measured in an EEG as a negative peak around 100ms after the stimulus presentation. This timing precisely matches the propagation of the fast transient component from the cortex to the SNr (∼70ms after the stimulus presentation according to the ICA algorithm) plus the transmission delay from the SNr to the cortex through the thalamus, which has been estimated around 35ms.[73–75] In any case, additional research with computational models (possibly including thalamic and cortical areas in the loop with BG) is required to better understand this process.

5. Conclusion

In this article we propose a new analysis method for evaluating transient phenomena, and it has been applied to the activity of BG populations in the framework of a detailed computational model. These novel metrics allow the explicit assessment of how cortical activity is transferred to the thalamus through the BG. We have analyzed how the relevant independent components of the signals in the input and output layers of the BG are affected with HD affectation and tonic DA levels. This combined study of DA and HD represents an innovative contribution, explaining the non-monotonic relationship between DA/HD levels and the selectivity of the BG. This paper describes the complex relations between BG

neuronal populations that are in accordance with the behavioral results that have been observed in the literature.

Acknowledgements

This research is supported by the University of Granada under FEDER 2014-2020, by the Andalucía Regional funds under the grants EmbBrain (A-TIC-276-UGR18) and CEREBIO (FEDER- P18-FR-2378) and National Grant (MICINN-FEDER-PID2019-109991GB-I00). This research has also received funding from the EU H2020 Framework Program under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3). Additionally, the main author has been funded with a national research training grant (FPU17/04432). Finally, the 3D character model used to illustrate this article is taken from Adobe's Mixamo platform.

Appendix A Neuron models

All the neurons included in our BG model have been simulated using different versions of the Izhikevich neuron model. This model is computationally very efficient and allows the reproduction of all the firing patterns previously described in the BG.[28,33] According to the Izhikevich model, the membrane potential v of the neuron is updated according to Eq. (A.1)

$$C\frac{dv}{dt} = k(v - v_r)(v - v_t) - u + I \tag{A.1}$$

where I is the total synaptic input (defined below), C is the membrane capacitance, v_r is the resting potential, v_t is the instantaneous threshold potential, k is an abstract parameter that regulates the influence of the current membrane potential value in its derivative and u is a recovery parameter updated by Eq. (A.2)

$$\frac{du}{dt} = a\,(b(v - v_r) - u) \tag{A.2}$$

where a sets the time scale of the recovery variable with low values corresponding to slow recoveries, and b describes the sensitivity of the recovery variable to fluctuations of the membrane potential.

An action potential is elicited in this model when the firing threshold (v_{peak}) is exceeded by the membrane potential v. In this case, the variables in the model are updated according to Eq. (A.3)

$$v \leftarrow c; \quad u \leftarrow u + d \tag{A.3}$$

where c is the voltage reset value and d is the reset of the recovery variable.

All the neuron sub-types defined in the MSN (D1 and D2), GPe (A, B and C) and SNr can be implemented using the original Izhikevich neuron model. On the contrary, the three neuron sub-types of the STN show different responses after long depolarization, including rebound bursts (RB), long-lasting rebound spikes (LLRS) and no rebound effect (NR). These effects have been modeled by extending the original Izhikevich's equations with one additional recovery variable (u_2).[62] The state variables are updated according to the following differential equations

$$C\frac{dv}{dt} = k(v - v_r)(v - v_t) - u_1 - w_2 \cdot u_2 + I \tag{A.4}$$

$$\frac{du_1}{dt} = a_1(b_1(v - v_r) - u_1) \tag{A.5}$$

$$\frac{du_2}{dt} = a_2(Gb_2(v - v_{r2}) - u_2) \tag{A.6}$$

where one additional recovery variable (u_2) and its parameters (a_1, a_2, b_1, b_2, d_1, d_2, w_1, w_2, G and U) have been added to account for the previously described behavior without losing the basic repertoire of firing patterns supported by the basic recovery variable u_1.[34,76] For the NR neurons, G is set to 1, while for RB and LLRS neurons, $G = H(v_{r2} - v)$ is the Heaviside step function:

$$H(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & x = 0 \\ 1 & x > 0 \end{cases} \tag{A.7}$$

When the membrane potential moves above the adaptive firing threshold ($v \geq v_{peak} + Uu_2$) the model variables are set as indicated in the following expressions:

$$v = c - Uu_2 \tag{A.8}$$

$$u_1 = u_1 + d_1 \tag{A.9}$$

$$u_2 = u_2 + d_2 \tag{A.10}$$

Finally, at high firing rates u_2 may increase dramatically. To avoid this phenomenon the U value is defined according to the following expression:

$$U = \frac{1}{w_1 \left| u_2 \right| + \frac{1}{w_1}} \tag{A.11}$$

The value of all these neuron model parameters for each cell type can be found in the tables 2 to 4 in https://github.com/EduardoRosLab/BG_selectivity/raw/master/parameters_tables.pdf.

Appendix B Synapse models

The input current (I) targeting a neuron is defined as follows:[29]

$$I = I_{AMPA} + I_{NMDA} B(v) + I_{GABA} \tag{B.1}$$

I_{AMPA}, I_{NMDA} and I_{GABA} are current inputs from AMPA, NMDA and GABA receptors, and $B(v)$ is a term that models the voltage-dependent magnesium plug in the NMDA receptors[46] as follows:

$$B(v) = \frac{1}{1 + \frac{[Mg^{2+}]_0}{3.57} e^{-0.062v}} \tag{B.2}$$

where $\left[Mg^{2+}\right]_0$ is the equilibrium concentration of magnesium ions. The input current of each channel z is defined as follows:

$$I_z = y_z (E_z - v) \tag{B.3}$$

where y_z is a exponentially-decaying conductance representing the contribution of receptor z to the membrane potential, E_z is the reversal potential of receptor z and v is the current membrane potential of the neuron.

The value of all these synaptic parameters can be found in the table 1 in https://github.com/EduardoRosLab/BG_selectivity/raw/master/parameters_tables.pdf.

Appendix C Dopaminergic modulation model

In the MSN, the overall in-vivo effect of the DA receptors D1 and D2 is that the stimulation of the D1 receptors increases neuron excitability, while the stimulation of the D2 receptors decrements the neuron firing,[77] as expressed in Eq. (C.1) and (C.2). There are also neuromodulatory effects implemented following Ref. 1, 29 and 27, where da represents the global level of DA in the system. This influences the D1 and D2 DA receptors according to the neuromodulatory factors β_1 and β_2, respectively. Eq. (C.3) models the D1-receptor mediated enhancement of the inward-rectifying potassium current. Eq. (C.4) models the enhancement of the L-type Ca2+ current. Finally, Eq. (C.5) models the increased sensitivity to injection current following D2 activation.

$$I_{NMDA} \leftarrow I_{NMDA}(1 + \beta_1 \cdot da) \tag{C.1}$$
$$I_{AMPA} \leftarrow I_{AMPA}(1 - \beta_2 \cdot da) \tag{C.2}$$
$$v_r \leftarrow v_r(1 + \beta_1 \cdot da) \tag{C.3}$$
$$d \leftarrow d(1 - \beta_2 \cdot da) \tag{C.4}$$
$$k \leftarrow k(1 - \beta_1 \cdot da) \tag{C.5}$$

GPe and STN neurons also show DA neuromodulatory effects on their synaptic receptors, which have been modeled as follows:

$$I_{AMPA} \leftarrow I_{AMPA}(1 - \beta_1 \cdot da) \tag{C.6}$$
$$I_{NMDA} \leftarrow I_{NMDA}(1 - \beta_1 \cdot da) \tag{C.7}$$
$$I_{GABA} \leftarrow I_{GABA}(1 - \beta_2 \cdot da) \tag{C.8}$$

The value of all these dopaminergic modulation parameters for each cell type can be found in the tables 2 to 4 in https://github.com/EduardoRosLab/BG_selectivity/raw/master/parameters_tables.pdf.

Bibliography

1. A. Tomkins, E. Vasilaki, C. Beste, K. Gurney and M. D. Humphries, Transient and steady-state selection in the striatal microcircuit, *Frontiers in Computational Neuroscience* **7**(January) (2014) p. 192.
2. O. Hikosaka, Y. Takikawa and R. Kawagoe, Role of the Basal Ganglia in the Control of Purposive Saccadic Eye Movements, *Physiol Rev* **80**(3) (2000) 953–978.
3. P. Redgrave, T. J. Prescott and K. Gurney, The basal ganglia: A vertebrate solution to the selection problem?, *Neuroscience* **89**(4) (1999) 1009–1023.
4. S. Shipp, The functional logic of corticostriatal connections, *Brain Structure and Function* **222** (mar 2017) 669–706.
5. M. D. Humphries, R. D. Stewart and K. N. Gurney, A Physiologically Plausible Model of Action Selection and Oscillatory Activity in the Basal Ganglia, *Journal of Neuroscience* **26**(50) (2006) 12921–12942.

6. K. Gurney, T. J. Prescott and P. Redgrave, A computational model of action selection in the basal ganglia. I. A new functional anatomy, *Biological Cybernetics* **84** (may 2001) 401–410.
7. K. Gurney, T. J. Prescott and P. Redgrave, A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour, *Biological Cybernetics* **84** (may 2001) 411–423.
8. S. M. Suryanarayana, J. H. Kotaleski, S. Grillner and K. N. Gurney, Roles for globus pallidus externa revealed in a computational model of action selection in the basal ganglia, *Neural Networks* **109** (jan 2019) 113–136.
9. P. Rihet, C. A. Possamaï, J. Micallef-Roll, O. Blin and T. Hasbroucq, Dopamine and human information processing: A reaction-time analysis of the effect of levodopa in healthy subjects, *Psychopharmacology* **163**(1) (2002) 62–67.
10. M. Nance, J. S. Paulsen, A. Rosenblatt and V. Wheelock, *A Physician's Guide to the Management of Huntington's Disease*, third edit edn. (Huntington's Disease Society of America, feb 2011).
11. M. F. Beal and R. J. Ferrante, Experimental therapeutics in transgenic mouse models of Huntington's disease, *Nature Reviews Neuroscience* **5** (may 2004) 373–384.
12. Y. Deng, R. Albin, J. Penney, A. Young, K. Anderson and A. Reiner, Differential loss of striatal projection systems in Huntington's disease: a quantitative immunohistochemical study, *Journal of Chemical Neuroanatomy* **27** (jun 2004) 143–164.
13. C. Loeb, G. L. A. Medica, C. Albano and M. Diseases, Levodopa and Huntington's chorea (1976) 958–961.
14. C. Beste, C. Saft, O. Gunturkun and M. Falkenstein, Increased Cognitive Functioning in Symptomatic Huntington's Disease As Revealed by Behavioral and Event-Related Potential Indices of Auditory Sensory Memory and Attention, *Journal of Neuroscience* **28** (nov 2008) 11695–11702.
15. J. Sorinas, M. D. Grima, J. M. Ferrandez and E. Fernandez, Identifying Suitable Brain Regions and Trial Size Segmentation for Positive/Negative Emotion Recognition, *International Journal of Neural Systems* **29** (mar 2019) p. 1850044.
16. J. Gomez-Pilar, J. Poza, A. Bachiller, C. Gómez, P. Núñez, A. Lubeiro, V. Molina and R. Hornero, Quantification of Graph Complexity Based on the Edge Weight Distribution Balance: Application to Brain Networks, *International Journal of Neural Systems* **28** (feb 2018) p. 1750032.
17. S. Ghosh-Dastidar and H. Adeli, Spiking neural networks, *International journal of neural systems* **19**(04) (2009) 295–308.
18. S. Ghosh-Dastidar and H. Adeli, Improved spiking neural networks for eeg classification and epilepsy and seizure detection, *Integrated Computer-Aided Engineering* **14**(3) (2007) 187–212.
19. S. Ghosh-Dastidar and H. Adeli, A new supervised learning algorithm for multiple spiking neural networks with application in epilepsy and seizure detection, *Neural networks* **22**(10) (2009) 1419–1431.
20. A. Geminiani, C. Casellato, A. Antonietti, E. D'Angelo and A. Pedrocchi, A multiple-plasticity spiking neural network embedded in a closed-loop control system to model cerebellar pathologies, *International journal of neural systems* **28**(05) (2018) p. 1750017.
21. G. Antunes, S. F. Faria da Silva and F. M. Simoes de Souza, Mirror neurons modeled through spike-timing-dependent plasticity are affected by channelopathies associated with autism spectrum disorder, *International journal of neural systems* **28**(05) (2018) p. 1750058.
22. A. Antonietti, J. Monaco, E. D'Angelo, A. Pedrocchi and C. Casellato, Dynamic redistribution of plasticity in a cerebellar spiking neural network reproducing an associative learning task perturbed by tms, *International journal of neural systems* **28**(09) (2018) p. 1850020.
23. C. Liu, J. Wang, Y. Y. Chen, B. Deng, X. L. Wei and H. Y. Li, Closed-loop control of the thalamocortical relay neuron's Parkinsonian state based on slow variable, *International Journal of Neural Systems* **23**(4) (2013) 1–14.
24. F. Su, J. Wang, B. Deng, X. L. Wei, Y. Y. Chen, C. Liu and H. Y. Li, Adaptive control of Parkinson's state based on a nonlinear computational model with unknown parameters, *International Journal of Neural Systems* **25**(1) (2015) 1–13.
25. T. C. Stewart, X. Choo, C. Eliasmith and Others, Dynamic Behaviour of a Spiking Model of Action Selection in the Basal Ganglia, *Proceedings of the 10th international conference on cognitive modeling* (2010) 235–240.
26. T. C. Stewart, T. Bekolay and C. Eliasmith, Learning to select actions with spiking neurons in the basal ganglia, *Frontiers in Neuroscience* **6**(JAN) (2012) 1–14.
27. J. T. Moyer, J. A. Wolf and L. H. Finkel, Effects of Dopaminergic Modulation on the Integrative Properties of the Ventral Striatal Medium Spiny Neuron, *Journal of Neurophysiology* **98** (dec 2007) 3731–3748.
28. M. Humphries, Capturing dopaminergic modulation and bimodal membrane behaviour of striatal medium spiny neurons in accurate, reduced models, *Frontiers in Computational Neuroscience* **3**(November) (2009) 1–16.
29. M. D. Humphries, R. Wood and K. Gurney, Dopamine-modulated dynamic cell assemblies generated by the GABAergic striatal microcircuit, *Neural Networks* **22**(8) (2009) 1174–1188.
30. Z. Fountas and M. Shanahan, The role of cortical oscillations in a spiking neural network model of the basal ganglia, *PLOS ONE* **12** (dec 2017) p.

e0189109.
31. K. N. Gurney, M. D. Humphries and P. Redgrave, A New Framework for Cortico-Striatal Plasticity: Behavioural Theory Meets In Vitro Data at the Reinforcement-Action Interface, *PLoS Biology* **13**(1) (2015).
32. S. Kunkel, A. Morrison, P. Weidel, J. M. Eppler, A. Sinha, W. Schenck, M. Schmidt, S. B. Vennemo, J. Jordan, A. Peyser, D. Plotnikov, S. Graber, T. Fardet, D. Terhorst, H. Mørk, G. Trensch, A. Seeholzer, R. Deepu, J. Hahne, I. Blundell, T. Ippen, J. Schuecker, H. Bos, S. Diaz, E. Hagen, S. Mahmoudian, C. Bachmann, M. E. Lepperød, O. Breitwieser, B. Golosio, H. Rothe, H. Setareh, M. Djurfeldt, T. Schumann, A. Shusharin, J. Garrido, E. B. Muller, A. Rao, J. H. Vieites and H. E. Plesser, Nest 2.12.0 (jan 2017).
33. E. M. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting* (The MIT Press, 2005).
34. Z. Fountas, Action selection in the rhythmic brain: The role of the basal ganglia and tremor, PhD Thesis, PhD thesis, Imperial College London (2016), pp. 224–228.
35. M. D. Humphries, R. Wood and K. Gurney, Reconstructing the Three-Dimensional GABAergic Microcircuit of the Striatum, *PLoS Computational Biology* **6**(November) (2010).
36. T. Koós and J. M. Tepper, Inhibitory control of neostriatal projection neurons by GABAergic interneurons, *Nature Neuroscience* **2**(5) (1999) 467–472.
37. D. F. English, O. Ibanez-Sandoval, E. Stark, F. Tecuapetla, G. Buzsáki, K. Deisseroth, J. M. Tepper and T. Koos, GABAergic circuits mediate the reinforcement-related signals of striatal cholinergic interneurons, *Nature Neuroscience* **15**(1) (2012) 123–130.
38. M. R. DeLong, Activity of basal ganglia neurons during movement, *Brain Research* **40** (may 1972) 127–135.
39. J. Bugaysen, M. Bronfeld, H. Tischler, I. Bar-Gad and A. Korngreen, Electrophysiological Characteristics of Globus Pallidus Neurons, *PLoS ONE* **5** (aug 2010) p. e12001.
40. N. E. Hallworth, C. J. Wilson and M. D. Bevan, Apamin-Sensitive Small Conductance Calcium-Activated Potassium Channels, through their Selective Coupling to Voltage-Gated Calcium Channels, Are Critical Determinants of the Precision, Pace, and Pattern of Action Potential Generation in Rat Subthalamic Nu, *The Journal of Neuroscience* **23** (aug 2003) 7525–7542.
41. H. Nakanishi, A. Tamura, K. Kawai and K. Yamamoto, Electrophysiological studies of rat substantia nigra neurons in an in vitro slice preparation after middle cerebral artery occlusion, *Neuroscience* **77** (feb 1997) 1021–1028.
42. F.-M. Zhou and C. Lee, Intrinsic and integrative properties of substantia nigra pars reticulata neurons, *Neuroscience* **198** (dec 2011) 69–94.
43. C. R. Lee and J. M. Tepper, Morphological and physiological properties of parvalbumin- and calretinin-containing γ-aminobutyric acidergic neurons in the substantia nigra, *The Journal of Comparative Neurology* **500** (feb 2007) 958–972.
44. M. Lindahl, I. Kamali Sarvestani, O. Ekeberg and J. H. Kotaleski, Signal enhancement in the output stage of the basal ganglia by synaptic short-term plasticity in the direct, indirect, and hyperdirect pathways., *Frontiers in computational neuroscience* **7**(June) (2013) p. 76.
45. H. Nakanishi, H. Kita and S. Kitai, Intracellular study of rat substantia nigra pars reticulata neurons in an in vitro slice preparation: electrical membrane properties and response characteristics to subthalamic stimulation, *Brain Research* **437** (dec 1987) 45–55.
46. C. E. Jahr and C. F. Stevens, Voltage Dependence of NMDA-Activated Predicted by Single-Channel Kinetics, *Journal of Neuroscience* **10**(9) (1990) 3178–3182.
47. M. M. Fan and L. A. Raymond, N-Methyl-d-aspartate (NMDA) receptor function and excitotoxicity in Huntington's disease, *Progress in Neurobiology* **81**(5-6) (2007) 272–293.
48. J. W. Goodliffe, H. Song, A. Rubakovic, W. Chang, M. Medalla, C. M. Weaver and J. I. Luebke, Differential changes to D1 and D2 medium spiny neurons in the 12-month-old Q175+/- mouse model of Huntington's Disease, *PLOS ONE* **13** (aug 2018) p. e0200626.
49. A. J. McGeorge and R. L. Faull, The organization of the projection from the cerebral cortex to the striatum in the rat, *Neuroscience* **29**(3) (1989) 503–537.
50. T. Moriizumi and T. Hattori, Pyramidal cells in rat temporoauditory cortex project to both striatum and inferior colliculus, *Brain Research Bulletin* **27**(1) (1991) 141–144.
51. X. Guo, H. Yu, N. X. Kodama, J. Wang and R. F. Galán, Fluctuation scaling of neuronal firing and bursting in spontaneously active brain circuits., *International journal of neural systems* **30**(1) (2020) p. 1950017.
52. G. Cantwell, M. Riesenhuber, J. L. Roeder and F. G. Ashby, Perceptual category learning and visual processing: An exercise in computational cognitive neuroscience, *Neural Networks* **89** (2017) 31–38.
53. M. J. Frank, By Carrot or by Stick: Cognitive Reinforcement Learning in Parkinsonism, *Science* **306** (dec 2004) 1940–1943.
54. H. Schroll, J. Vitay and F. H. Hamker, Working memory and response selection: A computational account of interactions among cortico-basalganglio-thalamic loops, *Neural Networks* **26** (2012) 59–74.
55. Z. Fountas and M. Shanahan, Assessing selectivity in the basal ganglia: The "gearbox" hypothesis, *bioRxiv*

(2017).
56. E. Oja and A. Hyva, Independent component analysis : algorithms and applications, **13** (2000) 411–430.
57. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* **12** (2011) 2825–2830.
58. R. Schmidt, D. K. Leventhal, N. Mallet, F. Chen and J. D. Berke, Canceling actions involves a race between basal ganglia pathways, *Nature neuroscience* **16**(8) (2013) p. 1118.
59. R. Natarajan and B. K. Yamamoto, The basal ganglia as a substrate for the multiple actions of amphetamines, *Basal Ganglia* **1** (jun 2011) 49–57.
60. B. R. Miller, A. G. Walker, A. S. Shah, S. J. Barton and G. V. Rebec, Dysregulated Information Processing by Medium Spiny Neurons in Striatum of Freely Behaving Mouse Models of Huntington's Disease, *Journal of Neurophysiology* **100** (oct 2008) 2205–2216.
61. S. Devarajan, P. S. Prashanth and V. S. Chakravarthy, The role of the basal ganglia in exploratory behavior in a model based on reinforcement learning, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **3316**(2) (2004) 70–77.
62. Z. Fountas and M. Shanahan, Phase offset between slow oscillatory cortical inputs influences competition in a model of the basal ganglia, *2014 International Joint Conference on Neural Networks (IJCNN)*, (IEEE, jul 2014), pp. 2407–2414.
63. C. Eliasmith, T. C. Stewart, X. Choo, T. Bekolay, T. DeWolf, C. Tang, D. Rasmussen, Y. Tang, C. Tang and D. Rasmussen, A Large-Scale Model of the Functioning Brain, *Science* **338**(6111) (2012) 1202–1205.
64. D. Caligiore, F. Mannella, M. A. Arbib and G. Baldassarre, Dysfunctions of the basal ganglia-cerebellar-thalamo-cortical system produce motor tics in Tourette syndrome, *PLOS Computational Biology* **13** (mar 2017) p. e1005395.
65. D. Caligiore, F. Mannella and G. Baldassarre, Different dopaminergic dysfunctions underlying Parkinsonian Akinesia and tremor, *Frontiers in Neuroscience* **13**(MAY) (2019) 1–15.
66. J.-P. Vonsattel, R. H. Myers, T. J. Stevens, R. J. Ferrante, E. D. Bird and E. P. Richardson, Neuropathological Classification of Huntington's Disease, *Journal of Neuropathology and Experimental Neurology* **44** (nov 1985) 559–577.
67. D. Caligiore and M. H. Fischer, Vision, action and language unified through embodiment, *Psychological Research* **77**(1) (2013) 1–6.
68. D. Caligiore, G. Pezzulo, G. Baldassarre, A. C. Bostan, P. L. Strick, K. Doya, R. C. Helmich, M. Dirkx, J. Houk, H. Jörntell, A. Lago-Rodriguez, J. M. Galea, R. C. Miall, T. Popa, A. Kishore, P. F. M. J. Verschure, R. Zucca and I. Herreros, Consensus Paper: Towards a Systems-Level View of Cerebellar Function: the Interplay Between Cerebellum, Basal Ganglia, and Cortex, *The Cerebellum* **16** (feb 2017) 203–229.
69. D. Caligiore, M. A. Arbib, R. C. Miall and G. Baldassarre, The super-learning hypothesis: Integrating learning processes across cortex, cerebellum and basal ganglia, *Neuroscience & Biobehavioral Reviews* (2019).
70. F. Mannella and G. Baldassarre, Selection of cortical dynamics for motor behaviour by the basal ganglia, *Biological Cybernetics* **109**(6) (2015) 575–595.
71. F. Li, W. Peng, Y. Jiang, L. Song, Y. Liao, C. Yi, L. Zhang, Y. Si, T. Zhang, F. Wang, R. Zhang, Y. Tian, Y. Zhang, D. Yao and P. Xu, The Dynamic Brain Networks of Motor Imagery: Time-Varying Causality Analysis of Scalp EEG, *International Journal of Neural Systems* **29** (feb 2019) p. 1850016.
72. V. B. Penhune and C. J. Steele, Parallel contributions of cerebellar, striatal and M1 mechanisms to motor sequence learning, *Behavioural Brain Research* **226**(2) (2012) 579–591.
73. P. Robinson, C. J. Rennie, D. Rowe and S. O'connor, Estimation of multiscale neurophysiologic parameters by electroencephalographic means, *Human brain mapping* **23**(1) (2004) 53–72.
74. J. A. Roberts and P. A. Robinson, Modeling absence seizure dynamics: implications for basic mechanisms and measurement of thalamocortical and corticothalamic latencies, *Journal of theoretical biology* **253**(1) (2008) 189–201.
75. S. J. van Albada and P. A. Robinson, Mean-field modeling of the basal ganglia-thalamocortical system. i: Firing rates in healthy and parkinsonian states, *Journal of theoretical biology* **257**(4) (2009) 642–663.
76. Z. Fountas and M. Shanahan, GPU-based fast parameter optimization for phenomenological spiking neural models, *Proceedings of the International Joint Conference on Neural Networks* **2015-Septe** (2015).
77. A. R. West and A. A. Grace, Opposite Influences of Endogenous Dopamine D 1 and D 2 Receptor Activation on Activity States and Electrophysiological Properties of Striatal Neurons: Studies Combining In Vivo Intracellular Recordings and Reverse Microdialysis, *The Journal of Neuroscience* **22** (jan 2002) 294–304.

frontiers

Black-box and surrogate optimization for tuning spiking neural models of striatum plasticity

https://doi.org/10.3389/fninf.2022.1017222

N.C. Cruz [1], **A. González-Redondo** [1,*], **J.L. Redondo** [2], **J.A. Garrido** [1], **E.M. Ortigosa** [1] **and P.M. Ortigosa** [2]

[1] *Dept. of Computer Engineering, Automation and Robotics, University of Granada, Granada, Spain*
[2] *Dept. of Informatics, University of Almería, ceiA3 Excellence Agri-food Campus, Almería, Spain*

Correspondence*:
A. González-Redondo
alvarogr@ugr.es

ABSTRACT

The basal ganglia (BG) is a brain structure that has long been proposed to play an essential role in action selection, and theoretical models of spiking neurons have tried to explain how the BG solves this problem. A recently proposed functional and biologically inspired network model of the striatum (an important nucleus of the BG) is based on spike-timing-dependent eligibility (STDE) and captures important experimental features of this nucleus. The model can recognize complex input patterns and consistently choose rewarded actions to respond to such sensory inputs. However, model tuning is challenging due to two main reasons. The first is the expert knowledge required, resulting in tedious and potentially biased trial-and-error procedures. The second is the computational cost of assessing model configurations (approximately 1.78 hours per evaluation). This work studies how to address the model tuning problem through numerical optimization. Considering the cost of assessing solutions, the selected methods stand out due to their low requirements for solution evaluations and compatibility with high-performance computing. They are the SurrogateOpt solver of Matlab and the RBFOpt library, both based on radial basis function approximations, and DIRECT-GL, an enhanced version of the widespread black-box optimizer DIRECT. Besides, a parallel random search serves as a baseline reference of the outcome of opting for sophisticated methods. SurrogateOpt turns out to be the best option for tuning this kind of model. It outperforms, on average, the quality of the configuration found by an expert and works significantly faster and autonomously. RBFOpt and the random search share the second position, but their average results are below the option found by hand. Finally, DIRECT-GL follows this line becoming the worst-performing method.

Keywords: striatum, reinforcement learning, spiking neural network, dopamine, spike-timing-dependent plasticity, model tuning, surrogate optimization, black-box optimization

1 INTRODUCTION

Computational models of the brain are useful tools for studying learning mechanisms. However, the difficulty involved in finding parameters that provide good solutions is a major challenge.

A model already published in a previous article (Gonzalez-Redondo et al., 2022) is a complex model that is difficult to obtain good solutions for. This model tries to better understand how learning through interaction to achieve a goal is solved by animals (or agents) by choosing among many possible actions to obtain rewards, as described in the reinforcement learning (RL) paradigm (Sutton et al., 1992). The model is based on spike-timing-dependent eligibility (Gurney et al., 2015) (STDE), a learning rule capturing important experimental features in the brain and, specifically the basal ganglia (BG, a set of nuclei located in the forebrain). This brain structure is related with the process of action-selection, according to biological (Grillner et al., 2005; Graybiel, 1998; Hikosaka et al., 2000) and computational studies (Gurney et al., 2001; Tomkins et al., 2014; Redgrave et al., 1999). We implemented (Gonzalez-Redondo et al., 2022) a functional and biologically inspired network model of the striatum (STR, an important input nucleus of the BG), where learning is based on STDE. The proposed model has been demonstrated to be capable of recognizing input patterns relevant to the task and consistently choosing rewarded actions in response to that input.

However, models require tuning (Martínez-Álvarez et al., 2016; Van Geit et al., 2007), and the quality expectations, datasets, and adaptability requirements are continuously growing (Masoli et al., 2020; Van Geit et al., 2008). The model described in (Gonzalez-Redondo et al., 2022), which attracts the attention of this work, contains dozens of free parameters: learning kernel shapes, synaptic and neuron time constants, lateral inhibition weight, etc. Some of them can be inferred from experimental data, but most of them must be manually tuned with plausible values. With this number of parameters, the curse of dimensionality leads to a tedious trial-and-error search procedure prone to failures. Another problem is the computational cost of evaluating each model configuration: it takes approximately 1.78 hours per evaluation in a modern laptop using a single CPU core. Both made the tuning of our model slow (the parameters finally used were found after two months of search), sub-optimal (as there are a huge parametric space not covered) and biased (by the intuition of the expert). Fortunately, model tuning can be addressed as a global optimization problem. There exist modern frameworks, such as Ray[Tune] (Liaw et al., 2018) and Vizier (Golovin et al., 2017), which implement multiple algorithms compatible with this purpose. Besides, the current increase in computer power that allows for defining more sophisticated models also helps us to face more challenging optimization problems (Cruz et al., 2021; Marín et al., 2021; Van Geit et al., 2008).

When addressing model tuning as an optimization problem, the objective function generally represents the difference between the desired and achieved output of the model for any candidate configuration. Concerning the associated problem, when the objective function exhibits mathematically exploitable properties, such as linearity, convexity, and continuous variables, it can be exactly solved. Otherwise, its resolution can be significantly challenging (Lindfield and Penny, 2017; Salhi, 2017). This issue might arise when the objective function does not have a closed analytical form or relies on sophisticated models with non-linear expressions, uncertainty, and simulations (Cruz et al., 2018; Marín et al., 2021). Luckily, some methods aim at finding acceptable results with a reasonable effort by using randomness and intuitive ideas. Most heuristics and meta-heuristics would fall into this group (Lindfield and Penny, 2017; Salhi, 2017). Similarly, if a method does not have specific knowledge or strict requirements for the objective function apart from being able to evaluate candidate solutions, it is classified as a black-box optimizer (Audet and Hare, 2017; Golovin et al., 2017). Both categories are frequently linked, as many meta-heuristics, such as evolutionary and swarm intelligence algorithms, are also black-box methods.

In this context, black-box optimization methods can be classified into two groups: those without specific components to require few function evaluations and those with them. It could be said that needing few function evaluations to converge is one of the goals pursued when designing any optimization method. However, most population-based meta-heuristics need numerous function evaluations (Costa and Nannicini, 2018) to compensate their instability due to randomness (Jones and Martins, 2021). They would hence fall into the first group. For instance, for the successful evolutionary optimizer UEGO (Cruz et al., 2018; García-Martínez et al., 2015; Marín et al., 2021), a robust configuration could need up to 1,000,000 function evaluations (Ortigosa et al., 2001). This potential requirement is usually attenuated with parallel computing, which fits well with population-based algorithms (Cruz et al., 2019; Jelásity, 2013; Storn and Price, 1997). This can be seen as a brute-force approach to tackle the high consumption of function evaluations. The methods in the second group do not renounce the benefits of high-performance computing, but they try to avoid function evaluations by design. Their use can be the only option when the cost of evaluating the objective function cannot be hidden with parallel computing. The most relevant methods in this group are surrogate optimizers (Bhosekar and Ierapetritou, 2018; Costa and Nannicini, 2018; Vu et al., 2017), which avoid evaluating the real objective function by constructing a lightweight model of it. They define an active research line in Global Optimization.

In this work, the objective function is not a plain mathematical function, such as a parabola. Instead, each evaluation launches a process that consists in building the neural network according to the input parameters of the candidate configuration, training it, and returning its performance at the target task. As mentioned above, this process is computationally demanding. For this reason, this work pays attention to optimization algorithms requiring few function evaluations. The selection consists of four solvers in total. The first two are SurrogateOpt, provided by the official Global Optimization Toolbox of Matlab (López, 2014), and RBFOpt (Costa and Nannicini, 2018), open-source and written in Python. Both construct a surrogate of the real objective function by combining radial basis ones (Gutmann, 2001; Regis and Shoemaker, 2007). The third method is DIRECT-GL (Stripinis et al., 2018), an enhanced version of the widespread DIRECT (Jones and Martins, 2021), which stands out due to its deterministic and effective strategy of dividing the search space and prioritizing the most promising areas to save function evaluations. The last one is a simple random search (Cruz et al., 2018), which is expected to define the baseline performance. Nonetheless, this method has also been implemented to benefit from parallel computing, so its rate of candidate solution evaluation is high. To the best of the authors' knowledge, the tuning of spiking neural models of striatum plasticity has not been studied from this perspective before. Hence, the ultimate goal of this research is to recommend the most effective strategy for this purpose to save tedious trial-and-error procedures and hyper-parameter tuning for Spiking Neural Networks (SNNs) by hand in general, which is inherently biased by the expert.

The rest of the paper is structured as follows: Section 2 describes materials and methods. Section 3 contains the experimentation and results. Finally, Section 4 shows the conclusions and states future work.

2 MATERIALS AND METHODS

This section starts with a detailed description of the computational models that define the agent behavior and the task it is solving. After that, the four optimization strategies considered are explained.

2.1 Computational models

For the network model we used conductance-based versions of the Leaky-Integrate and Fire (LIF) neuron model (Gerstner and Kistler, 2002). LIF model simplifies many aspects of neuronal dynamics, so it is

more computationally efficient than other commonly used neural models in SNNs. We use this model in every layer of the network. Before the use of optimization methods, the parameters were manually tuned to obtain reasonable firing rates (see details in Supplementary Materials). The STR neurons are divided in D1 or D2 populations, each one with different learning kernel constants (so they can learn to respond to different situations; more on this later), and also divided by channels (one per action). STR D1 and D2 populations are complementary, as the D1 population tries to learn what action it has to do, while D2 population tries to learn what action it has to stop. The action neurons are a population that integrates its channel activity and outputs the agent's behavior, and they are tuned to fire every input cycle if they receive enough stimulation (at least two more spikes from D1 neurons than D2 neurons each cycle). The dopamine neuron was tuned to have a firing range from 50 to 350 spikes per second, with these unrealistic values chosen to improve computational performance (instead of simulating a bigger dopaminergic population).

The input generation procedure is described in Gonzalez-Redondo et al. (2022) and based on Masquelier et al. (2009); Garrido et al. (2016). The agent perceives the environment as 2000 analog inputs. These inputs are fed one-to-one to an input layer of LIF neurons as currents (Figure 1B), altogether with an oscillatory drive. This oscillatory drive leads to a current-to-phase conversion: the neurons that receive the strongest analog input currents will fire first during the phase of the cycle (Masquelier et al., 2009). This way we encode analog inputs in specific spatio-temporal spike activity patterns. This is called phase-of-firing encoding and represents information in the spike times of neurons relative to the phase of a background oscillation (in our case, the oscillatory drive). New input stimuli are presented at uniformly distributed random intervals of 200-500ms. The stimulus can be a repeating pattern or noise, and both are generated randomly depending on the simulation seed. When presenting a repeating pattern, only half of the input neurons (1000) are pattern-specific, while the other half receives random current values. When no pattern is presented, all the input neurons receive random current values.

The network model (Figure 1A) contains two channels. Every channel contains two parallel layers (STR D1 and D2 neurons, respectively) of striatal-like neurons with asymmetrical structured lateral inhibition (as in Burke et al. (2017)) within and between STR D1 and D2 populations. The output of each channel is an action node that integrates the channel activity to decide if the agent takes an action or not. The agent can do none, both, or any of them at a time. A dopaminergic neuron projects its activity to both action channels as a neuromodulator (dopamine) determining what the agent should learn from the recent past experience. An environment reward signal (based on the chosen and the expected action) is delivered to this neuron as excitatory (rewards) or inhibitory (punishments) input.

The neurons in each channel receive plastic synapses from the input layer. The STDE (Gurney et al., 2015) learning rule is used, a modification of a reward-modulated STDP learning rule where the kernel constants are dopamine-dependent (that is, different values are defined for low dopamine and high dopamine values, see Figure 2). This rule also uses elegibility traces to store the potential changes, similarly to Izhikevich (2007). The learning kernels are different for STR D1 and D2 neurons, as their biological counterparts respond to different situations (Gerfen and Surmeier, 2011): D1 neurons are more predominant in the direct pathway of the BG, which tend to promote behavior when it is active. D2 neurons are more predominant in the indirect pathway of the BG, which tend to inhibit behavior when it is active. For this reason the initial learning kernels were manually chosen to be complementary: D1 neurons learn to do actions, and D2 neurons learn to stop actions. The dopaminergic modulatory signal is global and delivered to every STDE connection from input layer to channel neurons. Lastly, two homeostatic mechanisms are added to improve the learning process: first, the synapses implementing the STDE included a non-Hebbian strengthening in

response to every pre-synaptic spike. Second, we include adaptive threshold to our neuron models based on Galindo et al. (2020).

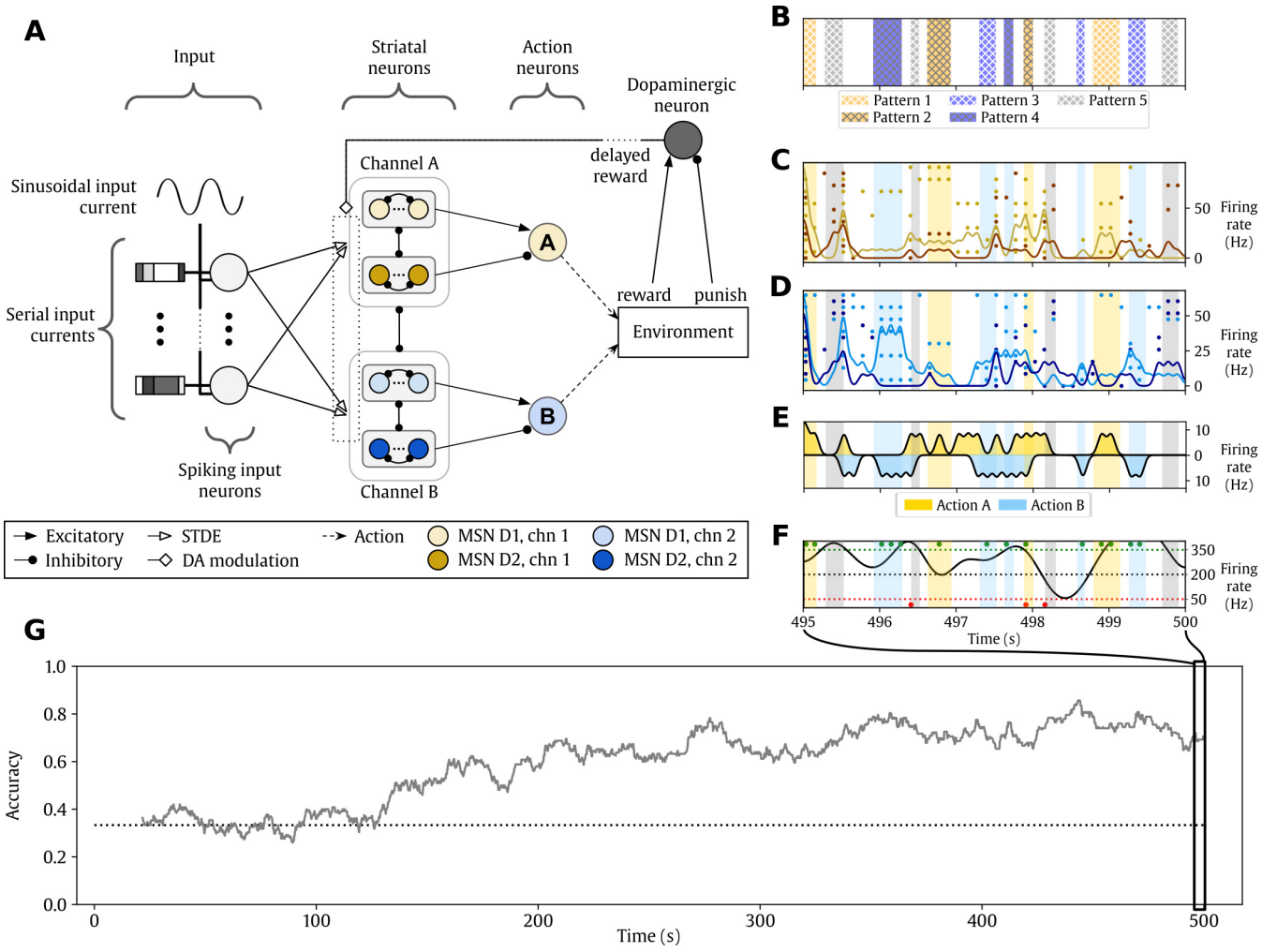

Figure 1. Cortico-striatal network solving a RL task (from Gonzalez-Redondo et al. (2022)). **A**. Structure of the network. **B-F**. The activity of the network during the last 5 seconds of simulation. Background color indicates the reward policy (yellowish colors, action A is rewarded and B is punished; bluish colors, action B is rewarded and A is punished; grey, any action is punished). **B**. Input pattern conveyed to the input layer. **C**. Raster plot of the channel-A action neurons. Yellow dots represent STR D1 spikes, and orange dots are STR D2 spikes. **D**. Raster plot of channel B. Cyan dots represent STR D1 spikes, and dark blue dots are STR D2 spikes. **E**. Action neuron firing rates. The middle horizontal line represents 0 Hz. Action A and B activity are represented in opposites directions for clarity. Action A neuronal activity increases in yellow zones while action B neuronal activity in cyan intervals. **F**. Firing rate of the dopaminergic neuron (black line). Dotted horizontal lines indicate the range of dopamine activity considered: black is the baseline, green is the maximum reward, and red represents the maximum punishment. Dots indicate rewards (green) and punishment (red) events delivered to the agent. **G**. Evolution of the learning accuracy of the agent, see Section 2.1 for further details. The dotted line marks the accuracy level by chance.

The agent has to learn a simple mapping task from stimulus to action. Every 200-500ms a new stimulus is presented and the agent has to respond with an appropriate action. There are five different repeating patterns and the agent has two possible actions to choose, A or B. Two patterns require action A, other two patterns require action B, and the fifth pattern requires do nothing. If the agent responds correctly, the

environment gives a reward. If a different action is taken, a punishment is given. If the input is just noise, the environment does not give rewards nor punishments.

We use a confusion matrix to help measure the performance of the model. Each row indicates the rewarded action in response to the presented pattern, and each column indicates the selected action in response to the presented pattern. Every cell m_{ij} then counts the number of occurrences of j action being done when i action was expected to be done. We only consider in the calculation those trials in which some reward or punishment can be delivered, ignoring those intervals with only noise as the stimulus. We consider that an action has been taken if the corresponding action neuron has spiked at least once during the pattern presentation, and conversely, we consider that no action has been taken if none of the action neurons spikes during the same duration. By doing so, we obtain a confusion matrix, widely used in classification problems when the objective is to describe the accuracy of a final map process (Stehman, 1997). The confusion matrix is defined as in expression (1),

$$C \equiv \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1C} \\ m_{21} & m_{22} & \cdots & m_{2C} \\ \vdots & \vdots & \ddots & \vdots \\ m_{C1} & m_{C2} & \cdots & m_{CC} \end{bmatrix} \tag{1}$$

where m_{ij} represents the number of occurrences belonging to the i-th class (the rewarded action) but classified as members of the j-th class (the selected action).

We then measure the model's performance as the accuracy of the classification, defined as the sum of the number of correct predictions (the trace of the matrix) divided by the total number of pattern presentations considered (the sum of the whole matrix). In order to measure the evolution through time of the performance of the models, we calculate the confusion matrix for each pattern presentation and then use a rolling mean of the last 100 values to obtain an estimation of the temporal evolution of the accuracy.

2.2 Model tuning as an optimization problem

In this context, it is possible to measure the performance of the model resulting from any set of parameters as the accuracy, F, of the classification, according to Equation (2). This value is defined as the sum of the number of correct predictions (the trace of the matrix) divided by the total number of pattern presentations considered (the sum of the whole matrix). To measure the evolution through time of the performance of the models, we calculate the confusion matrix for each pattern presentation and then use a rolling mean of the last 100 values to obtain an estimation of the temporal evolution of the accuracy.

$$F = \frac{\sum_i m_{ii}}{\sum_i \sum_j m_{ij}} \tag{2}$$

The value of F will ultimately depend on thirteen variables, a selected subset of all the variables of the model, which determine the model behavior. Notice that the model features inherent stochasticity, which is handled by returning the average of five simulations. The variables are shown in Table 1, including their corresponding ranges. These variables have been chosen to be optimized as they are the ones related to the learning process of the model. We did not optimize the neuron model variables as we already found reasonable values to make their firing behavior match their biological counterparts. Variable w_{max} represents the maximum weight of each plastic synapse. Variable C_{pre} is the homeostatic term applied per

presynaptic spike. Variable τ_{th} is the time constant of the adaptive neuron threshold. Variable C_{th} defines the additive increment of the adaptive threshold of a striatal neuron after a spike, and it is inversely proportional to the target firing rate. Variable μ is a dimensionless constant that modulates all learning parameters. Lastly, the kernel shape of the STDE learning rule is defined by the parameters k_{DA}^{SPK} with $SPK \in \{+, -\}$ being the spike order pre-post for applying k_{DA}^{+} and post-pre for applying k_{DA}^{-}, respectively, and $DA \in \{hi, lo\}$ being the high- or low-DA cases, resulting in four parameters per neuron population: k_{hi}^{+}, k_{lo}^{+}, k_{hi}^{-} and k_{lo}^{-}. As we have two neuron populations $POP \in \{d1, d2\}$, there are eight $k_{POP_{DA}}^{SPK}$ STDE parameters in total: $k_{d1_{hi}}^{+}$, $k_{d1_{lo}}^{+}$, $k_{d1_{hi}}^{-}$, $k_{d1_{lo}}^{-}$, $k_{d2_{hi}}^{+}$, $k_{d2_{lo}}^{+}$, $k_{d2_{hi}}^{-}$ and $k_{d2_{lo}}^{-}$. A graphical representation of all these kernels for the manually-tuned case can be found in Figure 2.

Table 1. Parameters to tune for the neural model and their allowed ranges.

Variable	Lower bound	Upper bound	Unit
w_{max}	10^{-3}	10^{-1}	μS
C_{pre}	-10^{-5}	10^{-5}	μS
μ	$5 \cdot 10^{-4}$	$5 \cdot 10^{-2}$	-
τ_{th}	1	200	seconds
C_{th}	10^{-2}	2	mV
$k_{d1_{hi}}^{+}$	-1	1	-
$k_{d1_{hi}}^{-}$	-1	1	-
$k_{d1_{lo}}^{+}$	-1	1	-
$k_{d1_{lo}}^{-}$	-1	1	-
$k_{d2_{hi}}^{+}$	-1	1	-
$k_{d2_{hi}}^{-}$	-1	1	-
$k_{d2_{lo}}^{+}$	-1	1	-
$k_{d2_{lo}}^{-}$	-1	1	-

Based on this quality metric and the variables involved, model tuning can be expressed as an optimization problem. It focuses on finding the values of the parameters (within their feasible range) that maximize the value of F, which becomes the objective function in optimization terms. The problem can be formulated according to Equation (3). For simplicity, only the first and the last variables are shown. The constraints keep every variable in its feasible range, which results in a box-constrained problem (Costa and Nannicini, 2018; Stripinis and Paulavičius, 2022). The max and min superscripts linked to each parameter symbol denote its upper and lower bounds, respectively. The numerical values are those shown in Table 1.

$$\begin{aligned} \underset{w_{max},\ldots,k_{d2_{lo}}^{-}}{\text{maximize}} \quad & F(w_{max}, \ldots, k_{d2_{lo}}^{-}) \\ \text{subject to} \quad & w_{max}^{lower} \leq w_{max} \leq w_{max}^{upper} \\ & \ldots \\ & k_{d2_{lo}}^{-,lower} \leq k_{d2_{lo}}^{-} \leq k_{d2_{lo}}^{-,upper} \end{aligned} \tag{3}$$

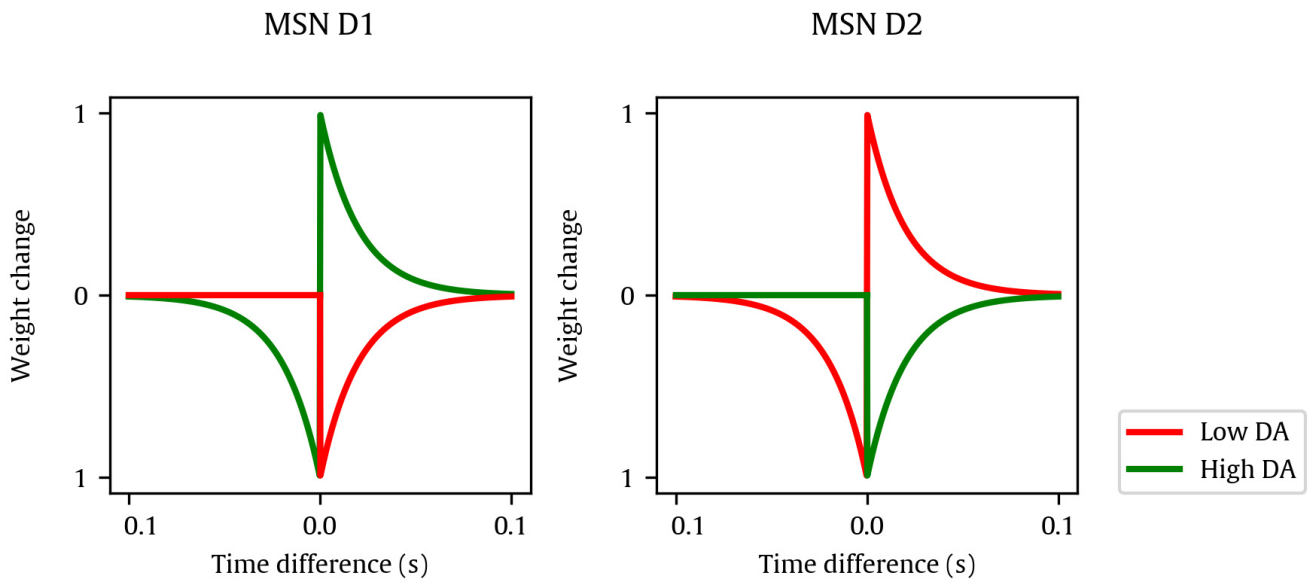

Figure 2. Manually-tuned kernels used for STDE synapses of MSN D1 (left) and D2 (right), showing the weight change depending on the time difference between pre- and post-synaptic spikes and dopamine. Lines represent kernels at dopamine minimum and maximum values (red and green, respectively).

Notice that the problem is defined as a maximization one, but optimization methods traditionally aim at minimization. Regardless, this is not relevant because converting a maximization problem into a minimization one is trivial. It is only necessary to multiply the objective function by -1, i.e., maximizing $F(\dots)$ is equal to minimizing $-F(\dots)$.

2.3 Optimization methods

As introduced, evaluating the objective function relies on non-deterministic simulations and is computationally demanding. Thus, the methods considered are designed for black-box optimization (Audet and Hare, 2017), i.e., RBFOpt (Costa and Nannicini, 2018), SurrogateOpt (Matlab, 2021), DIRECT-GL (Stripinis et al., 2018), and a random search Cruz et al. (2018). The first three, which are also the preferred options, have been explicitly designed to require few function evaluations. All of them are prepared for exploiting parallel computing. Finally, it is relevant to highlight that among the considered methods, DIRECT-GL is the only deterministic one, which means that the algorithm does not rely on randomness and always returns the same result for the same problem instance and configuration.

2.3.1 RBFOpt

RBFOpt, published in Costa and Nannicini (2018), is an open-source library written in Python for black-box optimization with computationally-expensive objective functions. This tool is based on the method proposed by Gutmann (2001).

RBFOpt belongs to the family of surrogate optimization methods. The fundamental idea of surrogate optimization is that the process relies on iteratively building an approximate model (response surface or surrogate model) of the real objective function. While the former approximates the latter, its computational requirements are expected to be significantly lower, and the accuracy can improve as the information on the target function increases with the points evaluated (Vu et al., 2017). For building the surrogate

Table 2. Frequent radial basis functions.

$\phi(r)$	**Type**	**Minimum degree**
r	Linear	0
r^3	Cubic	1
$r^2 \log r$	Thin plate spline	1

model, RBFOpt uses radial basis functions, whose output depends on the distance between the input and a given reference. In this field, Gutmann (2001) was a pioneer of using radial basis functions for optimizing computationally demanding black-box functions (Costa and Nannicini, 2018).

Let $f(x)$ be an abstract objective function of the form $f : \mathbb{R}^N \to \mathbb{R}$, where $x \in \left[x^{min}, x^{max}\right]$, and $x^{min}, x^{max} \in \mathbb{R}^N$, i.e., the corresponding lower and upper bounds of each decision variable. Notice that this can be seen as a generalization of the particular problem formulation in Equation (3). For K different points of the search space, $x_1, \ldots, x_K$, with known values, $y_1 = f(x_1), \ldots, y_K = f(x_K)$, the associated radial basis function interpolant, s_K, has the following structure as the sum of K radial basis functions (Costa and Nannicini, 2018; Vu et al., 2017):

$$s_K(x) = \sum_{i=1}^{K} \lambda_i \phi\left(\|x - x_i\|\right) + p(x) \tag{4}$$

where $\phi : \mathbb{R}_+ \to \mathbb{R}$, which is a radial basis function, $\lambda_1, \ldots, \lambda_K \in \mathbb{R}$ acting like weights of the model, and $p(x)$ is a polynomial. The minimum degree of p to guarantee the existence of the interpolant depends on the form of ϕ. Table 2 contains three common radial basis functions for a generic input, r. It also includes the minimum degree of their accompanying polynomial p that ensures the existence of the interpolant. If these components are appropriately configured, the desired radial basis function interpolant can be efficiently computed by solving a linear system to find the unknown parameters, such as the weights. For instance, $x_1, \ldots, x_K$ should be pairwise distinct (Costa and Nannicini, 2018; Vu et al., 2017).

The general procedure applied by optimization methods using radial basis functions follows Algorithm 1 (Costa and Nannicini, 2018). A particular method will define a strategy to implement these generic steps, starting from selecting the initial points. For example, for low-dimensional problems, a valid approach is to choose the corners of the search space. Another one is to pick the corners and the central point. Controlling the effort put into improving the accuracy of the surrogate model and finding the best point with the current model is also critical. The method by Gutmann (2001) defined a measure of the bumpiness of the surrogate model for this purpose. Their method assumes that the real objective function does not oscillate excessively, so when configuring models and considering new points, the smoother (or 'least bumpy') interpolant is preferred (see Figure 3, which assumes four known points and a hypothetical target value of the cost function). Regardless, describing these aspects in detail is out of the scope of this paper. See the work by Vu et al. (2017) to have a detailed overview, and that by Costa and Nannicini (2018) to understand the fundamentals of RBFOpt.

In this context, RBFOpt has two main contributions. The first is an automatic model selection component. The second is the support for using faster yet less accurate variants of the objective function. The latter is especially appropriate for the target problem since the simulation-related parts of the objective function, such as the training time and the seeds, are adjustable. They can be modified by the expert in charge of

Algorithm 1 Generic global optimization through radial basis functions

```
Initial step - Select K points
while There is available function evaluations do
    Compute the radial basis function interpolant
    Decide between improving the surrogate model and finding the best point using the current model.
    Determine the next point to consider according to the previous decision
    Evaluate the objective function at the new point
end while
return Best point found
```

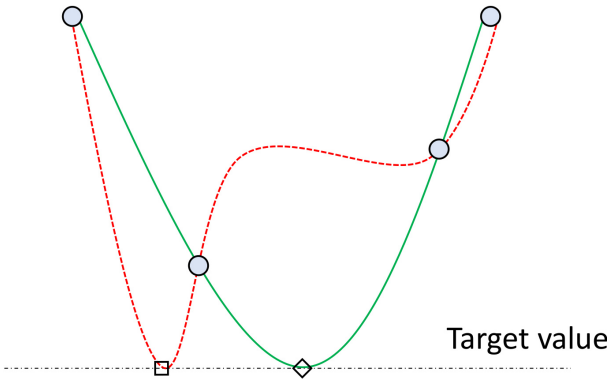

Figure 3. Depiction of two surrogate models interpolating four points (blue circles) and reaching a target value (horizontal dashed line). The green solid-line model is considered more likely than the red dashed-line one since it is smoother. In other words, the method by Gutmann (2001) assumes that it is more likely that the point tagged with a diamond exists (green line) rather than that with a square (red line) in the real function (Costa and Nannicini, 2018).

model tuning to reduce time at the expense of losing accuracy. These properties, along with its open-source nature, the compatibility with parallel computing, and the good results reported in Costa and Nannicini (2018) motivated its consideration for the present work.

2.3.2 SurrogateOpt

SurrogateOpt is a solver for computationally-demanding black-box optimization problems provided by the Global Optimization Toolbox (López, 2014) of Matlab Matlab (2021) since its version R2018b. As introduced, it belongs to the same group as RBFOpt since the method is a surrogate optimization algorithm. SurrogateOpt also uses radial basis function interpolators. Its documentation motivates this decision by highlighting that they support any number of dimensions and are computationally cheap to construct, evaluate, and extend. This tool is mainly based on the algorithm proposed by Regis and Shoemaker (2007). It has been selected due to its effectiveness, simplicity of use, and compatibility with parallel computing.

Conceptually, SurrogateOpt follows a scheme similar to Algorithm 1. The fundamental differences correspond to the implementation of each step and specific definitions. In this regard, SurrogateOpt has a rich set of associated concepts and procedures, which are summarized below. According to its official documentation, the method alternates between two stages: Construct Surrogate and Search for Minimum. The change between them occurs after what is called the surrogate reset.

In the first stage, the method builds a surrogate of the real objective function. For this purpose, it interpolates a radial basis function through a set of points whose value must be computed with the real yet computationally demanding objective function. SurrogateOpt uses a cubic radial basis function with

a linear tail, which minimizes the concept of bumpiness (Gutmann, 2001) previously mentioned when describing RBFOpt. In the beginning, the solver computes and evaluates a user-given number of random points distributed adequately within the bounds. It can also start from a user-given set of points of known value. In later executions of this stage, the software package will create and evaluate a parameter-defined number of random points. As explained for RBFOpt, building the desired interpolant involves solving a linear system of equations.

In the second stage, SurrogateOpt looks for a minimum of the objective function using a procedure that resembles a local search. More specifically, the method defines a search region radius, known as the scale, whose initial value is 0.2. It starts from the best point since the last surrogate reset, i.e., the one with the smallest objective function value. This point is called the incumbent point. The search then focuses on finding a minimum of a merit function that relates the surrogate and the distance from the points evaluated with the real objective function. This approach aims to find a trade-off between minimizing the surrogate, which is not the real objective function and is potentially less accurate, and evaluating new points accurately.

Mathematically, the definition of the merit function for any point x combines two weighted terms, the scaled surrogate, $S(x)$, and the scaled distance, $D(x)$. Being s_{min} and s_{max} the minimum and maximum surrogate values of the sample points, respectively, and $s(x)$ that of the considered point, the scaled surrogate is defined as follows (Matlab, 2021):

$$S(x) = \frac{s(x) - s_{min}}{s_{max} - s_{min}} \tag{5}$$

$S(x)$ is non-negative and zero at points having minimal surrogate values among sample points. Concerning the scaled distance, it is defined as follows:

$$D(x) = \frac{d_{max} - d(x)}{d_{max} - d_{min}} \tag{6}$$

where d_{min} and d_{max} are the minimum and maximum distances from a sample point to any evaluated one, respectively, and $d(x)$ is the minimum distance of the point x to an evaluated one. $D(x)$ is non-negative, and zero at points at the furthest distance from evaluated points. Hence, minimizing $D(x)$ orientates the algorithm towards regions separated from evaluated points. The merit function is a convex combination of both parts according to the following structure:

$$wS(x) + (1 - w)D(x) \tag{7}$$

where w is a weighting factor between zero and one. The greater it is, the most effort is put into minimizing the surrogate model. Analogously, the smaller it is, the most interest in exploring new regions. This weighting factor cycles through the following values, according to Regis and Shoemaker (2007): 0.3, 0.5, 0.8, and 0.95.

During the search, the solver adds multiple (up to thousands) random vectors to the incumbent point to generate sample points. The vectors are shifted and scaled by the bounds in each dimension and ultimately multiplied by the scale. The sample points must also respect the problem bounds. Then, the merit function is evaluated at all of them further than a parameter-defined distance from any point previously evaluated. The one featuring the best (lowest) value of the merit function becomes an adaptive point. The real objective function will be ultimately computed at it, which will be used to update the surrogate model and assess the

real gain from the incumbent value. If the real value of the adaptive point is significantly better than the current incumbent point, the former replaces the latter, and the search is considered successful. Otherwise, the incumbent point remains unaltered, and the search is classified as unsuccessful.

The scale of the search changes when one of the following conditions is met:

1. There have been three successful searches since the last scale change.
2. There have been either five or the number of problem variables (whichever greater) unsuccessful searches since the last scale change.

If the first condition is met, the scale is doubled (up to a maximum length of 0.8 times the size of the box defined by the problem bounds). If the second situation occurs first, the scale is divided by two (without becoming lower than $1e{-}5$ times the size of the box defined by the problem bounds). By proceeding this way, the search ultimately focuses near an incumbent point featuring a small objective function value.

After considering all the new sample points further than a minimum distance from evaluated points, the Search for Minimum phase ends to go back to the Construct Surrogate one, i.e., resetting the surrogate model. This phase change generally occurs after reducing the scale until all sample points are closely around the incumbent point.

2.3.3 DIRECT-GL

DIRECT-GL, proposed by Stripinis et al. (2018); Stripinis and Paulavičius (2022), is an enhanced version of a popular method, DIRECT (Jones and Martins, 2021). This new variant is designed as a modification of a specific part of the original method. Hence, it is convenient to start by describing the initial DIRECT and its framework, inherited by the new one.

DIRECT was proposed in Jones et al. (1993) as a modification of Lipschitzian Optimization that did not require specifying a Lipschitz constant, i.e., a bound on the rate of change of the objective function, which cannot be easily computed in real problems (or it may not exist). Aside from keeping a deterministic behavior, the method was simpler, converged faster, and featured a certain degree of compatibility with parallel computing. It was later revised by his author in Jones (2001) to handle not only box or domain constraints and continuous variables, but also nonlinear inequality constraints and integer variables. From the beginning, this method was conceived for black-box optimization and situations in which the objective function was time-consuming.

Global optimization algorithms must find a trade-off between exploration and exploitation of the search space (Van Geit et al., 2008). The first term refers to finding unexplored regions, and the second represents the capacity to find the best solution in a known zone (global and local search capabilities, respectively). In Lipschitz Optimization, the Lipschitz constant is treated as a weighting factor determining how much emphasis to put into global over local search by indicating where to split the search space into sub-regions. This value must equal or exceed the maximum rate of change of the objective function, so conservative configurations excessively focus on global search. It also makes these methods slow to converge because modifying the value at search is challenging. In contrast to them, DIRECT could maintain the scheme of dividing the search space and autonomously prioritized the regions to explore by virtually considering all possible constants. The division was also independent of the number of dimensions, so the algorithm was more scalable with the problem dimensionality (yet not recommended for more than 20 variables (Jones, 2001)).

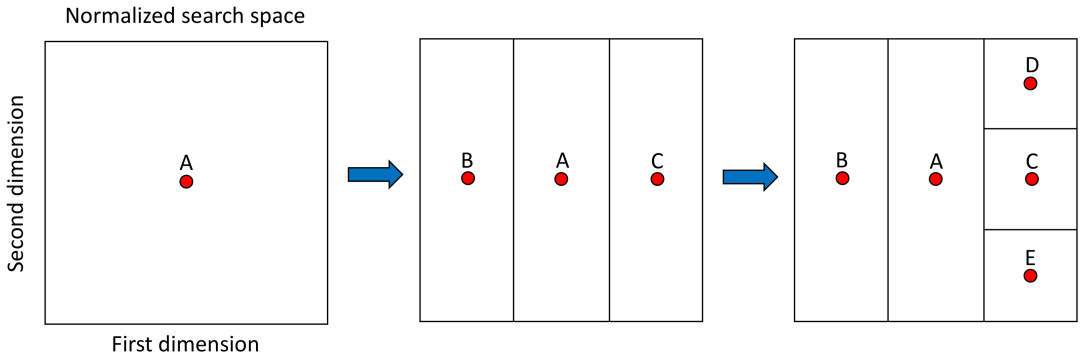

Figure 4. Main aspects of DIRECT and its search space after two iterations. At the beginning of the initial iteration, point A is the first evaluated and represents the first and only rectangle (square) defined by the search space. It is trisected and results in the rectangles defined by points B and C. At the second iteration, the rectangle defined by point C is selected and trisected resulting in rectangles defined by points D and E.

More specifically, DIRECT starts by normalizing each variable to [0, 1] so that the search space becomes the unit hypercube. Then, the method proceeds by dividing it into sub-rectangles. This scheme determines the name of the method, since DIRECT comes from 'DIviding RECTangles'. The rectangles are represented by the value of the objective function at their center, which avoids the effect of problem dimensionality: rectangles only have one center independently of the dimensions. It is also relevant to highlight that the referred division is a trisection in reality, which allows keeping the focus on the original rectangle without further re-evaluation, i.e., its center stills refer to a different region. Figure 4 depicts these ideas assuming a 2D search space and two divisions (trisections).

The fundamental aspect of DIRECT is how the rectangles are selected for division and further exploration at each iteration. This selection is deterministic and theoretically considers every possible balance between exploration and exploitation (Lipschitz-like constant). As detailed in Jones (2001), a pure global method would always select the widest rectangle. A pure local one would opt for the one with the best value at its center. The former avoids overlooking the promising regions, while the latter promotes that.

DIRECT does not force itself to select just one rectangle, which would require parameters to tune. Instead, the method computes all the weightings of local versus global search. For this purpose, it defines the size of any rectangle as the distance between its center and one of its vertices. Then, for every selection at a particular iteration, the method represents all the available rectangles depending on their size and the value of their center. After that, it proceeds to select those in the low-right convex hull. Figure 5 depicts this idea assuming a minimization problem. The selected rectangles are the balanced options between local and global search considering the central value and size of the corresponding regions. Notice the similitude of this approach to computing the Pareto set as the solution to a multi-objective optimization problem (Filatovas et al., 2017).

Interestingly, as explained in Jones (2001) and also shown in Figure 5, the selection of rectangles can be alternatively derived from the rate of change of the function in each. If one knows the optimal value, anchors a half-line at it, and swings the free extreme upwards, the first dot touched represents the rectangle with the most reasonable rate of change, i.e., gradual instead of steep, to contain the optimum. Hence, that rectangle must be selected. In reality, the optimal value is not usually known. However, it is possible to repeat this process from the best value known so far, as the optimal value will be equal to or lower than it, to minus infinity. Selecting the touched dot for each anchored point results in the lower-right convex hull previously defined. It is hence possible to obtain the same selection scheme yet by thinking differently.

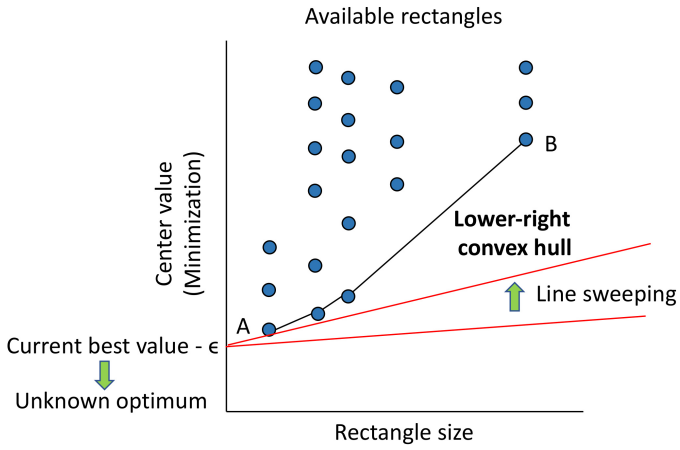

Figure 5. Selection of rectangles to further explore (divide) in DIRECT for a hypothetical minimization problem. The available rectangles are represented as dots (circles are shown for clarity). The horizontal axis corresponds to the size of each rectangle, and the vertical one shows the value of the objective function at its center (lower is better for minimization). In this context, rectangle A is the one with the best central value, while B encompasses the broadest region. The rectangles selected will lie in the lower-right convex hull, which represents the optimal balance between exploration and exploitation. The red lines at the bottom shows an alternative way to derive this selection method: A line is anchored at every value better than known, starting from that one minus the desired accuracy (minimum relevant change), ϵ, and swept upwards until reaching a rectangle. Repeating the process to (negative for minimization) infinity results in the same convex hull (and avoids regions with expected negligible improvements).

Besides, it is possible to subtract an arbitrary value, ϵ, to the best value known to discard from the hull the rectangles with negligible improvements. Accordingly, DIRECT only expects as input the maximum number of function evaluations and the constant ϵ, which can be seen as the desired accuracy of the solution. The interested reader can see Jones (2001) for further information about this algorithm.

Despite its good properties (conceptual simplicity, ingenious deterministic exploration, and requiring a single parameter), DIRECT is not free of potential drawbacks, and researchers have proposed numerous variants (Jones and Martins, 2021; Stripinis and Paulavičius, 2022). The two main flaws of the initial method are (Stripinis and Paulavičius, 2022) i) the potential waste of function evaluations in sub-optimal regions for functions with many local optima and ii) the slow convergence rate even after having identified the basin of the global optimum (Jones and Martins, 2021). Accordingly, the method selected for this work is one of the revised versions of DIRECT, i.e., DIRECT-GL, which tries to overcome both (Stripinis et al., 2018).

For this purpose, the authors of DIRECT-GL modified the selection of rectangles to consider more than its ancestor. The process has two stages and fits into the original framework without requiring extra parameters. The first one enhances the global search component of the method, represented by the letter ‘G’ in its name. It starts by adding the rectangles with the best central value and prioritizing those that are bigger. This approach results in more rectangles of medium size and the best values. The second phase is similar, but it considers the Euclidean distance to the best point known so far instead of the objective function value. Namely, it tries to add more hyper-rectangles close to the current minimum. This strategy strengthens the exploration of the most promising area, i.e., it enhances the local search aspect of the method, represented by the letter ‘L’ in its name.

Aside from the computational studies in Stripinis et al. (2018), the effectiveness of this strategy is supported by the recent comparison in Stripinis and Paulavičius (2022), where DIRECT-GL exhibits the

best performance among all the DIRECT-based methods. The implementation in Stripinis and Paulavičius (2022), also used in this work, unifies the results of both stages in a single selection. This aspect differs from the original work to make the method more suitable for parallelization and more effective.

2.3.4 Random Search

A pure random search procedure is arguably the simplest global optimizer (Brooks, 1958), and it belongs to the stochastic family of optimization methods (Cruz et al., 2018). More specifically, it consists in randomly generating solutions in the search space while keeping a record of the best one found so far. Algorithm 2 describes this process in detail. Notice that it is expressed in general terms, and the comparison criterion will depend on if the objective function is to be minimized or maximized.

Algorithm 2 Random search

Require: Objective function: f, Evaluations allowed: $evals$
1: solution $\leftarrow \emptyset$
2: currentVal $\leftarrow$ worst value
3: iter $\leftarrow 0$
4: **while** iter $<$ evals **do**
5: point $\leftarrow$ random()
6: **if** f(point) is better than currentVal **then**
7: solution $\leftarrow$ point
8: currentVal $\leftarrow f$(point)
9: **end if**
10: iter $\leftarrow$ iter $+ 1$
11: **end while**
12: **return** solution

Despite its simplicity, this method converges to a global optimum when the number of allowed evaluations tends to infinity (Brooks, 1958). On the one hand, its practical applicability is low due to the lack of orientation during the search. For this reason, it has been initially selected for the problem at hand as the expected baseline reference, especially considering that the computational cost of the objective function makes it difficult to work with high evaluation budgets. Thus, the previous methods are expected to outperform this one because of their sophisticated components to explore and exploit the search space (Van Geit et al., 2008). On the other hand, the simple structure of this procedure, which is also embarrassingly parallel in terms of high-performance computing (Trobec et al., 2018), ensures a high rate of solution evaluations in an appropriate computing platform. Hence, its results can be of interest depending on the ultimate problem difficulty and the quality requirements.

3 EXPERIMENTATION AND RESULTS

3.1 Problem-specific setup and reference value

A sample tuning problem of a spiking neural model of striatum plasticity has been addressed to assess the performance of the considered optimization methods. The problem was selected due to its high number of parameters, biological relevance, and computational cost of evaluating solutions. The model details are in (Gonzalez-Redondo et al., 2022), and the important problem-specific setup is summarized below.

The simulation time was 500 seconds, enough to the hand-tuned models to converge to a solution. The model contains 2000 leaky integrate-and-fire (LIF) input neurons and 16 spiking LIF output neurons with adaptive threshold divided in two channels (one per possible action). During the learning protocol 5

different repeating stimuli were used, besides noise. The duration of each stimulus is taken from a uniform random distribution between 100 and 500ms. Five different random seeds were used for every set of parameters tested and the resulting fitness of each seed averaged.

The best result obtained without optimization methods is shown in Figure 1. Panels C-F show network activity and rewards/punishments during the last 5 seconds of simulation. The most relevant information is the accuracy through the training process. The mean of the last 100 seconds of the accuracy is used as the fitness for the objective function. The procedure to calculate the accuracy is described in section 2.1. The accuracy evolution of the best result obtained by an expert after manually tuning for two months the parameters using a trial-and-error procedure is shown in Figure 1G. Good sets of parameters typically plateau after 400 seconds, as wrong actions are taken from time to time even with further training.

3.2 Computational setup

The computational platform used belongs to the high-performance computing cluster of the Supercomputing – Algorithms research group from the University of Almería, Spain. Specifically, up to 8 Bull Sequana X440-A5 nodes were used to launch different executions. Every node features 2 AMD EPYC Rome 7642 with 48 cores each, i.e., 96 cores in total, 512 GB of RAM, and 240 GB SSD as its main disk. In a core of one of these nodes, evaluating a candidate solution or set of model parameters for the configuration described above takes 1.78 ± 0.12 hours on average. This value has been computed by generating and evaluating 96 random feasible parameter sets. The software environment consists of Matlab 2020b for SurrogateOpt, DIRECT-GL, and Random Search, and Python 3.6.8 for RBFOpt.

Two computational budgets have been considered, 300 and 600 function evaluations. The first results from taking into account the estimated run time as follows: Since each evaluation takes 1.78 hours on average using a high-end processor, 300 evaluations should take $300 \times 1.78 = 534$ hours at least, i.e., 22 days approximately. This estimation assumes a sequential execution workflow, as a human expert will likely proceed, and neglects the overhead associated with the internal computations of the optimizers. The value of 22 days of work is in the same order of magnitude as the most favorable conditions found by the authors of this paper when doing the referred model fitting by hand. Similarly, the value of 300 function evaluations is equal to the budget used by RBFOpt by default. Concerning the second limit used, 600, it has been adjusted to two times the lower value. By proceeding this way, it will be possible to assess the benefits of doubling the effort. This value would also be close to the maximum function evaluations that Surrogateopt would assign to the problem at hand, namely, $50 \times 13 = 650$ according to the official documentation.

The sequential run time estimations would be 22 and 44 days, approximately. The former, for 300 evaluations, is demanding, but the latter, for 600, starts to be overwhelming for a person. Nevertheless, when automating the process by using optimizers compatible with parallel computing and there is access to a cluster, both conditions met in this work, the run time can be significantly lower. Ideally, by deploying 96 threads, the objective function evaluation time could be reduced by a factor up to 96. This speedup would mean turning the 534 hours turn into 5.56, approximately. In general, perfect speedup is achieved rarely. Spawning and managing concurrent execution units comes at a cost, sequential tasks do not benefit from them, and there might not always be enough work for all (e.g., few hyper-rectangles selected by DIRECT-GL at a particular iteration). Regardless, the time taken by the optimizers in the cluster is expected to be significantly lower than estimated above for a sequential execution.

Apart from controlling the number of function evaluations allowed, the four solvers have been configured with their default options. This includes configuring RBFOpt to use the Bonmin (Bonami et al., 2008)

and Ipopt (Wächter and Biegler, 2006) solvers (Costa and Nannicini, 2018) for addressing the internal sub-problems that arise (e.g., adjusting the radial basis function interpolants). Aside from this, notice that RBFOpt stands out by being capable of using a less accurate yet faster version of the objective function. Working with it requires both the referred kind of function and the lower and upper bounds of the expected error. To accelerate the neural model assessment, i.e., the objective function, the number of simulation seeds has been reduced from 5 to 1, which should make its computation five times faster on average.

The inaccuracy estimation has been computed as follows: 8 cluster nodes with the same specifications as defined above were used to generate 26880 feasible configurations randomly. Then, the standard deviation of the objective value for each one of the five simulation seeds was recorded. The average standard deviation between seeds for the same configuration was approximately 0.03. Then, according to the empirical rule of Statistics, this average standard deviation was multiplied by 3 to cover 99.7% of the values assuming a standard distribution. The result is 0.09, which was ultimately rounded up to 0.10 to add an arbitrary extra margin. Accordingly, the inaccurate yet faster function is passed to RBFOpt considering that the real value (if 5 simulations seeds were considered instead of 1) will be in the range of ± 0.10 plus the inaccurate estimation.

3.3 Numerical results

Table 3 contains the results for the model tuning problem addressed with each optimizer and function evaluation budget. The first column shows the optimization algorithm. The second one displays the number of function evaluations allowed. The values generally refer to the standard function with five simulation seeds. However, the two last cases of RBFOpt combine the full function with the one featuring a single simulation seed to be faster despite reducing its accuracy. They include the word 'fast' to highlight this aspect. Notice that dividing the fast term by 5 and adding it to the standard one results in the same budgets considered, i.e., 300 and 600 standard function evaluations. In the beginning, the second configuration of this type for RBFOpt consisted of 400 complete evaluations and up to 1000 fast ones, but the results were worst, and the solver opted for not executing that many fast evaluations. Thus, it seems preferable to put more emphasis on complete ones even though the estimated cost is theoretically equivalent, and we ultimately chose the configuration shown.

The following two columns contain the average efficiency (higher is better, with the best value in bold font) and the standard deviation for each optimizer and configuration. All the stochastic methods have been independently executed 20 times. With this information, the 95% confidence intervals have been computed according to the t-Student distribution considering the sample sizes (i.e., under 30 records each). They are shown in the fifth column. The sixth and last column contains the average run time for each case (the standard deviations are omitted because of not being either significant or especially relevant for this variable). For DIRECT-GL, the run times have been obtained by launching 8 independent executions, one per available node. Finally, it is relevant to mention that the RandomSearch results have been obtained from the dataset with 26880 random points used to assess the accuracy of the fast version of the objective function. More specifically, they come from taking 20 random samples with as many instances as the function evaluation budget. Thus, the run times of this method have been analytically estimated by multiplying the average evaluation time by the computational budget. They have been ultimately divided by the number of CPU cores due to the embarrassingly parallel nature of the process.

Concerning the results, the most noticeable aspect is that DIRECT-GL shows the worst performance in terms of achieved fitness and required run time. The aptitude of its solutions for 300 and 600 function evaluations is even worse than that obtained with the simplest method, i.e., RandomSearch. Both results,

Table 3. Performance metrics for each optimizer and configuration considered computed with the results of 20 independent executions.

Optimizer	Function evaluations	Average fitness	Standard Deviation	Confidence Interval (95%)	Average run time (h)
SurrogateOpt	300	0.7269	0.0667	[0.6957, 0.7581]	6.07
	600	**0.7699**	0.0691	[0.7376, 0.8022]	11.26
RBFOpt	300	0.5325	0.1461	[0.4641, 0.6009]	6.85
	600	0.6267	0.1434	[0.5596, 0.6938]	13.27
	200 + 500 fast	0.5843	0.1458	[0.5161, 0.6525]	5.30
	500 + 500 fast	0.5923	0.1550	[0.5198, 0.6648]	12.29
DIRECT-GL	300	0.4165	-	[0.4165]	81.98
	600	0.4618	-	[0.4618]	103.38
Random Search	300	0.5492	0.1412	[0.4831, 0.6153]	5.56
	600	0.6159	0.1027	[0.5678, 0.6640]	11.13

i.e., 0.4165 and 0.4618, stay outside of the confidence interval of this stochastic method, and below the lower bounds for 300 and 600 function evaluations. The same occurs when considering RBFOpt and its configuration with 300 evaluations. Accordingly, the difference between these methods is statistically significant. Its average run time is also significantly higher than the rest, which comes from the fact that the parallelism of DIRECT-GL is strictly bounded by the number of selected rectangles at any point. For this reason, it will not always exploit all the available CPU cores, which is critical in the context of interest.

Conversely, SurrogateOpt stands out as the best-performing method in terms of achieved fitness and a low standard deviation. The lower bound of its lowest confidence interval does not fall into the range of any other one, so the observed difference between these optimization strategies for the target problem is significant. Besides, considering that the run time of RandomSearch is an optimistic approximation, it could be said that the computational performance is virtually equivalent. Accordingly, the direct conclusion that can be drawn from the results shown in Table 3 is that SurrogateOpt is the best solver for this kind of model tuning problem. Moreover, its average results are comparable to that obtained by an expert after a tedious and time-demanding model tuning process. More specifically, as detailed in Section 3.1, the fitness of the expert-based model tuning was 0.7216, while the average of SurrogateOpt is 0.7269 with 300 function evaluations only. Nevertheless, one can doubt the effectiveness of doubling the computational budget for SurrogateOpt because the confidence intervals of both cases overlap.

When confidence intervals do not overlap, the difference between two groups is statistically significant. However, when they do, the difference might still be relevant Goldstein and Healy (1995); Sullivan (2008). To avoid this uncertainty, the confidence intervals for their difference will be computed. For this purpose, as both samples have less than 30 instances, the t-Student distribution will be used again. Notice that the following formulation assumes similar variances in the population, as it occurs between both cases of SurrogateOpt. The pooled estimate of the common standard deviation, S_P is computed as follows:

$$S_P = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}} \tag{8}$$

where n_1 and n_2 are the sample sizes of groups 1 and 2, respectively, and σ_1 and σ_2 refer to their corresponding standard deviation. With this information, the confidence interval for the difference between two means, $\overline{x}_1$ and $\overline{x}_2$, is obtained as follows:

$$(\overline{x}_1 - \overline{x}_2) \pm tS_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{9}$$

where t refers to the appropriate value from the t-Student table (determined by the sample sizes, as introduced) for the desired confidence level and $n_1 + n_2 - 2$ degrees of freedom. Notice that the two terms after t define the standard error of the difference in means between $\overline{x}_1$ and $\overline{x}_2$.

Back to the mean and standard deviations of both configurations of SurrogateOpt, the 95% confidence interval of their difference is [-0.0865, 0.0005] according to 9. It defines a range of likely values for the difference in means between both cases, $\overline{x}_1$ - $\overline{x}_2$, where $\overline{x}_1$ is that of 300 function evaluations and $\overline{x}_2$ is that of 600. Theoretically, since the interval contains the null value, i.e., 0, it can be concluded that there is no statistically significant difference between the average results of launching SurrogateOpt with 300 and 600 function evaluations. Even if the upper bound of this interval were slightly below zero, the average difference could be perceived as negligible yet confirmed. That said, in practical terms, since the outcome of this process is a model parameter set with the highest possible fitness, it seems reasonable to work with 600 function evaluations and several independent runs whenever possible.

Regarding RBFOpt and RandomSearch, they remain between the best performing solver, SurrogateOpt, and the worst, DIRECT-GL. Being free to use without cost, unlike SurrogateOpt, is their main attribute in this context. It is hard to find the best option among them at a glance due to the significant overlap and comparatively high standard deviations. Technically, RBFOpt offers the best average results when allowed to execute 600 function evaluations. Besides, the 95% confidence interval of the difference between 300 and 600 function evaluations confirms the effectiveness of doubling the computational effort, yet it is arguably negligible. However, the difference with the cases using fast evaluations is not statistically significant. The same occurs when comparing RandomSearch and RBFOpt with 600 function evaluations. This close similarity indirectly benefits RandomSearch, as it is the simplest strategy to apply, especially when there is an available parallel computing platform.

3.4 Insight into optimization-based model tuning results

The results of the optimization process yielded better learning capabilities than those of manual tuning. Figure 6A shows one of the best configurations found by SurrogateOpt, the preferred method, in action, and compares it to the manually tunned option. Although the manually tuned result plateaus after 400 seconds, the optimized result continues to improve accuracy after that point. In addition, the optimized result is more reliable, as its standard deviation is smaller.

Of the various parameters, the most interesting are the ones that define the shape of the STDE kernels of the learning rule for the neurons (MSN D1 and D2). Figure 6B shows the differences between the manually tuned kernel and the optimized kernel. While the manually tuned solution tends to use asymmetrical kernels in every case, it seems that the optimized solution uses symmetrical kernels for low DA and asymmetrical kernels for high DA.

If we consider symmetrical kernels having values with equal signs and asymmetrical kernels with opposite signs, this is in accordance with the values obtained by Gurney et al. (2015) in their exhaustive parametric search (Figure 11 in their article). This could be relevant as they are considering more biological constraints

than us. The only discrepancy is in the case of high DA in MSN D2 neurons, where the optimized kernel is reversed from the range obtained by Gurney et al. However, further research is needed to better understand the significance of these findings as well as the plausibility of the proposed parameters.

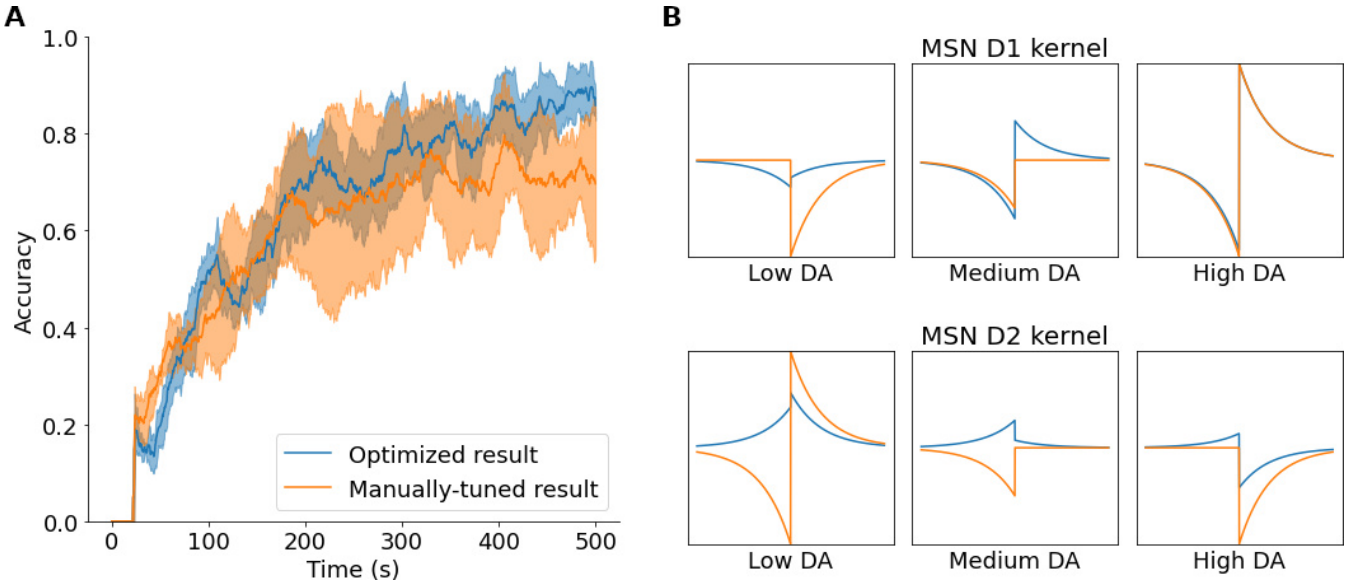

Figure 6. A. Comparison between one of the best results obtained with the SurrogateOpt optimization method (blue) with the best manually-tuned result (orange), with the mean and standard deviation ($n = 5$). **B.** Comparison of parameters related with the STDE kernel.

4 CONCLUSIONS AND FUTURE WORK

This study addresses the tuning of spiking neural models of striatum plasticity by using state-of-the-art black-box and surrogate optimization methods. This kind of model is useful for understanding how the brain could perform online reinforcement learning, a fundamental ability that is essential for many tasks such as motor control or decision-making. However, tuning these models is a difficult task due to the high dimensionality of the parameter space and the time required for the simulations. In addition, experts are often biased in their choices of parameter values. This problem can be addressed as an optimization problem, which can be solved using different methods.

This work makes a selection of optimization algorithms designed for computationally-demanding objective functions and compatible with parallel computing. The goal is to find the best alternative to avoid the necessity of tedious and expert-biased trial-and-error tuning of biologically realistic neural models that require much time to be simulated. This approach could automate model tuning despite not having been broadly studied yet in this context. Automation will not only avoid potential errors and biased configurations, but it can also reduce the required time from weeks to hours.

The solvers considered are SurrogateOpt, shipped with the Optimization Toolbox of Matlab, RBFOpt, which is an open-source optimizer written in Python, and DIRECT-GL, which is an improvement of a widespread optimizer written in Matlab yet open-source too. Aside from them, a naive pure-random search strategy has also been implemented. They have been compared when trying to tune a spiking neural model of striatum plasticity that takes 1.78 hours on average to be simulated in the computing platform. The methods were only allowed to evaluate 300 solutions in the first case and 600 in the second. Both

computational budgets are in the same order of magnitude as a human expert takes to tune the model used as the benchmark, i.e., several hundred function evaluations.

SurrogateOpt stands out as the best solver to use, and it is hence recommended for this kind of computationally demanding neural model tuning problem. It achieves the best average results with the lowest standard deviation and significantly distinguishes itself from the rest. The model configurations that it finds with 300 function evaluations can compete with the expert-based reference. Namely, the fitness of the expert-based model tuning was 0.7216 after two months of work, and the average of SurrogateOpt is 0.7269 with 300 function evaluations (6 hours approximately). This average increases up to 0.7699 when the method can launch 600 evaluations. However, the effectiveness of doubling the computational effort could not be confirmed on average for the studied problem. Regardless, the generic recommendation made is to work with the highest computational budget and multiple independent executions due to its stochastic nature.

RBFOpt and RandomSearch, both stochastic methods too, perform significantly worse than SurrogateOpt in terms of average fitness despite spending similar times. Hence, they should be only used when there is no access to the referred solver. Nevertheless, the potential of RandomSearch for this kind of problem is remarkable, especially when a high-performance computing platform is available. This method is trivial to implement, and its performance can be significantly improved by increasing the number of evaluated solutions per unit of time.

In contrast to the rest, DIRECT-GL, the only deterministic solver chosen, is also the worst option for the problem at hand. Its parallel computing capabilities are limited by the number of promising regions that the method can find. Since it does not find the best regions in the search space and finds few attractive zones, the algorithm is unable to fully exploit the computing platform. These aspects make it not only the solver that achieves the worst tuning configurations but also the slowest one.

As future work, the best-performing solver will be used to tune other neural models featuring computationally demanding simulation processes. Additionally, the study might be extended as new suitable methods arise.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

NC and AG conceived and designed the study. AG developed the neural model and manually tuned it, while NC linked the model to the optimization methods selected. JA, AG, and EM did the literature review concerning neural model tuning. NC, JL, and PM did the literature review regarding optimization algorithms. AG, JA, and EM studied the results in terms of Neuroinformatics and checked the quality of the optimization-based configurations. NC, JL, and PM analyzed and discussed the performance of each optimizer. NC and AG wrote the first version of the manuscript, and JA, JL, EM, and PM revised it and suggested improvements. All the authors made relevant contributions to the article and approved the submitted version.

FUNDING

This research has been funded by the R+D+i projects RTI2018-095993-B-I00 and PID2021-123278OB-I00, financed by MCIN/AEI/10.13039/501100011033/ and ERDF "A way to make Europe"; by the Junta de Andalucía with reference P18-RT-1193 and by the University of Almería with reference UAL18-TIC-A020-B. N.C. Cruz is supported by the Ministry of Economic Transformation, Industry, Knowledge and Universities from the Andalusian government.

ACKNOWLEDGEMENTS

This research is supported by the Spanish Grant INTSENSO (MICINN-FEDER-PID2019-109991GB-I00), Regional grants Junta Andalucía-FEDER (CEREBIO P18-FR-2378). This research has also received funding from the EU Horizon 2020 Framework Program under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3). Finally, A. González-Redondo is supported by an FPU Fellowship from the Spanish Ministry of Education (FPU17/04432).

DATA AVAILABILITY STATEMENT

The neural model used, the datasets generated, the cost functions implemented, and the optimization scripts defined will be available upon acceptance. *For now, the reviewers can access this material in a zip file attached during the upload.*

REFERENCES

Audet, C. and Hare, W. (2017). *Derivative-free and blackbox optimization*, vol. 2 (Springer)

Bhosekar, A. and Ierapetritou, M. (2018). Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Computers & Chemical Engineering* 108, 250–267

Bonami, P., Biegler, L. T., Conn, A. R., Cornuéjols, G., Grossmann, I. E., Laird, C. D., et al. (2008). An algorithmic framework for convex mixed integer nonlinear programs. *Discrete Optimization* 5, 186–204

Brooks, S. H. (1958). A discussion of random methods for seeking maxima. *Operations research* 6, 244–251

Burke, D. A., Rotstein, H. G., and Álvarez, V. A. (2017). Striatal local circuitry: a new framework for lateral inhibition. *Neuron* 96, 267–284

Costa, A. and Nannicini, G. (2018). Rbfopt: an open-source library for black-box optimization with costly function evaluations. *Mathematical Programming Computation* 10, 597–629

Cruz, N. C., Marín, M., Redondo, J. L., Ortigosa, E. M., and Ortigosa, P. M. (2021). A comparative study of stochastic optimizers for fitting neuron models. application to the cerebellar granule cell. *Informatica* 32, 477–498

Cruz, N. C., Redondo, J. L., Álvarez, J. D., Berenguel, M., and Ortigosa, P. M. (2018). Optimizing the heliostat field layout by applying stochastic population-based algorithms. *Informatica* 29, 21–39

Cruz, N. C., Salhi, S., Redondo, J. L., Álvarez, J. D., Berenguel, M., and Ortigosa, P. M. (2019). Design of a parallel genetic algorithm for continuous and pattern-free heliostat field optimization. *Journal of Supercomputing* 75, 1268–1283

Filatovas, E., Lančinskas, A., Kurasova, O., and Žilinskas, J. (2017). A preference-based multi-objective evolutionary algorithm R-NSGA-II with stochastic local search. *Central European Journal of Operations Research* 25, 859–878

Galindo, S. E., Toharia, P., Robles, O. D., Ros, E., Pastor, L., and Garrido, J. A. (2020). Simulation, visualization and analysis tools for pattern recognition assessment with spiking neuronal networks. *Neurocomputing* 400, 309–321. doi:10.1016/j.neucom.2020.02.114

García-Martínez, J. M., Garzón, E. M., Cecilia, J. M., Pérez-Sánchez, H., and Ortigosa, P. M. (2015). An efficient approach for solving the hp protein folding problem based on uego. *Journal of Mathematical Chemistry* 53, 794–806

Garrido, J. A., Luque, N. R., Tolu, S., and D'Angelo, E. (2016). Oscillation-Driven Spike-Timing Dependent Plasticity Allows Multiple Overlapping Pattern Recognition in Inhibitory Interneuron Networks. *International Journal of Neural Systems* 26, 1650020. doi:10.1142/S0129065716500209

Gerfen, C. R. and Surmeier, D. J. (2011). Modulation of striatal projection systems by dopamine. *Annual review of neuroscience* 34, 441

Gerstner, W. and Kistler, W. M. (2002). *Spiking neuron models: Single neurons, populations, plasticity* (Cambridge University Press)

Goldstein, H. and Healy, M. J. R. (1995). The graphical presentation of a collection of means. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 158, 175–177

Golovin, D., Solnik, B., Moitra, S., Kochanski, G., Karro, J., and Sculley, D. (2017). Google Vizier: A service for black-box optimization. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 1487–1495

Gonzalez-Redondo, A., Garrido, J., Naveros Arrabal, F., Hellgren Kotaleski, J., Grillner, S., and Ros, E. (2022). Reinforcement learning in a spiking neural model of striatum plasticity. Under review

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory* 70, 119–136. doi:https://doi.org/10.1006/nlme.1998.3843

Grillner, S., Hellgren, J., Ménard, A., Saitoh, K., and Wikström, M. A. (2005). Mechanisms for selection of basic motor programs – roles for the striatum and pallidum. *Trends in Neurosciences* 28, 364–370. doi:https://doi.org/10.1016/j.tins.2005.05.004

Gurney, K., Humphries, M. D., and Redgrave, P. (2015). A New Framework for Cortico-Striatal Plasticity: Behavioural Theory Meets In Vitro Data at the Reinforcement-Action Interface. *PLoS Biology* 13, e1002034. doi:10.1371/journal.pbio.1002034

Gurney, K., Prescott, T. J., and Redgrave, P. (2001). A computational model of action selection in the basal ganglia. i. a new functional anatomy. *Biological Cybernetics* 84, 401–410

Gutmann, H. M. (2001). A radial basis function method for global optimization. *Journal of global optimization* 19, 201–227

Hikosaka, O., Takikawa, Y., and Kawagoe, R. (2000). Role of the Basal Ganglia in the Control of Purposive Saccadic Eye Movements. *Physiological Reviews* 80, 953–978. doi:10.1152/physrev.2000.80.3.953

Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex* 17, 2443–2452. doi:10.1093/cercor/bhl152

Jelásity, M. (2013). Two approaches for parallelizing the uego algorithm. *Optimization Theory: Recent Developments from Mátraháza* 59, 159

Jones, D. R. (2001). The DIRECT global optimization algorithm. In *Encyclopedia of Optimization*, eds. C. A. Floudas and P. M. Pardalos (Boston: Springer). 431–440

Jones, D. R. and Martins, J. R. (2021). The direct algorithm: 25 years later. *Journal of Global Optimization* 79, 521–566

Jones, D. R., Perttunen, C. D., and Stuckman, B. E. (1993). Lipschitzian optimization without the lipschitz constant. *Journal of optimization Theory and Applications* 79, 157–181

Liaw, R., Liang, E., Nishihara, R., Moritz, P., González, J. E., and Stoica, I. (2018). Tune: A research platform for distributed model selection and training. *arXiv preprint arXiv:1807.05118*

Lindfield, G. and Penny, J. (2017). *Introduction to nature-inspired optimization* (Academic Press)

López, C. P. (2014). Optimization techniques via the optimization toolbox. In *MATLAB Optimization Techniques* (Springer). 85–108

Marín, M., Cruz, N. C., Ortigosa, E. M., Sáez-Lara, M. J., Garrido, J. A., and Carrillo, R. R. (2021). On the use of a multimodal optimizer for fitting neuron models. application to the cerebellar granule cell. *Frontiers in Neuroinformatics* 15

Martínez-Álvarez, A., Crespo-Cano, R., Díaz-Tahoces, A., Cuenca-Asensi, S., Ferrández Vicente, J. M., and Fernández, E. (2016). Automatic tuning of a retina model for a cortical visual neuroprosthesis using a multi-objective optimization genetic algorithm. *International Journal of Neural Systems* 26, 1650021

Masoli, S., Tognolina, M., Laforenza, U., Moccia, F., and D'Angelo, E. (2020). Parameter tuning differentiates granule cell subtypes enriching transmission properties at the cerebellum input stage. *Communications Biology* 3, 1–12

Masquelier, T., Hugues, E., Deco, G., and Thorpe, S. J. (2009). Oscillations, Phase-of-Firing Coding, and Spike Timing-Dependent Plasticity: An Efficient Learning Scheme. *Journal of Neuroscience* 29, 13484–13493. doi:10.1523/JNEUROSCI.2207-09.2009

Matlab, G. O. T. (2021). User's guide (r2021b). *The MathWorks Inc*

Miller, B. R., Walker, A. G., Shah, A. S., Barton, S. J., and Rebec, G. V. (2008). Dysregulated information processing by medium spiny neurons in striatum of freely behaving mouse models of huntington's disease. *Journal of neurophysiology* 100, 2205–2216

Ortigosa, P., García, I., and Jelasity, M. (2001). Reliability and performance of uego, a clustering-based global optimizer. *Journal of Global Optimization* 19, 265–289

Redgrave, P., Prescott, T. J. J., and Gurney, K. (1999). The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience* 89, 1009–1023. doi:10.1016/S0306-4522(98)00319-4

Regis, R. G. and Shoemaker, C. A. (2007). A stochastic radial basis function method for the global optimization of expensive functions. *INFORMS Journal on Computing* 19, 497–509

Salhi, S. (2017). *Heuristic search: The emerging science of problem solving* (Springer)

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment* 62, 77–89. doi:https://doi.org/10.1016/S0034-4257(97)00083-7

Storn, R. and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization* 11, 341–359

[Dataset] Stripinis, L. and Paulavičius, R. (2022). DGO: A new DIRECT-type MATLAB toolbox for derivative-free global optimization

Stripinis, L., Paulavičius, R., and Žilinskas, J. (2018). Improved scheme for selection of potentially optimal hyper-rectangles in direct. *Optimization Letters* 12, 1699–1712

Sullivan, L. M. (2008). *Essentials of biostatistics workbook: statistical computing using Excel* (Jones & Bartlett Learning)

Sutton, R. S., Barto, A. G., and Williams, R. J. (1992). Reinforcement learning is direct adaptive optimal control. *IEEE Control Systems Magazine* 12, 19–22

Tomkins, A., Vasilaki, E., Beste, C., Gurney, K., and Humphries, M. D. (2014). Transient and steady-state selection in the striatal microcircuit. *Frontiers in Computational Neuroscience* 7, 192. doi:10.3389/fncom.2013.00192

Trobec, R., Slivnik, B., Bulić, P., and Robič, B. (2018). *Introduction to parallel computing: from algorithms to programming on state-of-the-art platforms* (Springer)

765 Van Geit, W., Achard, P., and De Schutter, E. (2007). Neurofitter: a parameter tuning package for a wide
766 range of electrophysiological neuron models. *Frontiers in Neuroinformatics* 1, 1
767 Van Geit, W., De Schutter, E., and Achard, P. (2008). Automated neuron model optimization techniques: a
768 review. *Biological Cybernetics* 99, 241–251
769 Vu, K. K., d'Ambrosio, C., Hamadi, Y., and Liberti, L. (2017). Surrogate-based methods for black-box
770 optimization. *International Transactions in Operational Research* 24, 393–424
771 Wächter, A. and Biegler, L. T. (2006). On the implementation of an interior-point filter line-search
772 algorithm for large-scale nonlinear programming. *Mathematical Programming* 106, 25–57

SUPPLEMENTARY MATERIALS

773 Neuron models and parameters used

774 We used conductance-based versions of the Leaky-Integrate and Fire (LIF) neuron model (Gerstner
775 and Kistler, 2002) in every layer of the network, but with different parameters. We classify the neuron
776 types according to the layer they belong to: cortical neurons for the input, striatal neurons for the learning
777 layer, and action neurons for the output. There is also a dopaminergic neuron that receives the rewards and
778 punishments.

779 The parameters used for each type were manually tuned to obtain reasonable firing rates. For the cortical
780 neurons we used a number of spikes per input cycle (with 8 cycles per second) close to Masquelier et al.
781 (2009) and Garrido et al. (2016). For the striatal neurons, we tuned the parameters to obtain a mean firing
782 rate of around one spike per second to be within biological ranges (Miller et al., 2008) but with activity
783 peaks of two or three spikes per input cycle (16-24 spikes per second). The action neurons are tuned to fire
784 every input cycle if they receive enough stimulation from its channel (at least two more spikes from D1
785 neurons than D2 neurons each cycle). The dopamine neuron was tuned to have a firing range from 50 to
786 350 spikes per second. The parameters used for each neuron type are shown at supplementary Table 4.

Parameter	Cortical	Striatal	Action	Dopaminergic
e_{exc} (mV)	0.0	0.0	0.0	0.0
e_{inh} (mV)	-85.0	-85.0	-85.0	-85.0
τ_{AMPA} (ms)	5.0	5.0	5.0	5.0
τ_{GABA} (ms)	10.0	30.0	60.0	10.0
τ_{ref} (ms)	1.0	15.0	15.0	1.0
C_m (pF)	250.0	50.0	100.0	250.0
g_{leak} (nS)	25.0	10.0	25.0	25.0
V_{thr} (mV)	-40.0	-50.0	-40.0	-65.0
e_{leak} (mV)	-65.0	-65.0	-65.0	-40.0

Table 4. Neuron parameters used in the model.

Reinforcement Learning in a Spiking Neural Model of Striatum Plasticity

https://doi.org/10.1016/j.neucom.2023.126377

Álvaro González-Redondo[a,*], Jesús Garrido[a], Francisco Naveros Arrabal[a], Jeanette Hellgren Kotaleski[b], Sten Grillner[b], Eduardo Ros[a]

[a] *Research Centre for Information and Communications Technologies (CITIC-UGR). Calle Periodista Rafael Gómez Montero 2, E18071 Granada, Spain.*
[b] *Kungliga Tekniska Högskolan, SE-100 44, Stockholm, Sweden*

Abstract

The basal ganglia (BG), and more specifically the striatum, have long been proposed to play an essential role in action-selection based on a reinforcement learning (RL) paradigm. However, some recent findings, such as striatal spike-timing-dependent plasticity (STDP) or striatal lateral connectivity, require further research and modelling as their respective roles are still not well understood. Theoretical models of spiking neurons with homeostatic mechanisms, lateral connectivity, and reward-modulated STDP have demonstrated a remarkable capability to learn sensorial patterns that statistically correlate with a rewarding signal. In this article, we implement a functional and biologically inspired network model of the striatum, where learning is based on a previously proposed learning rule called spike-timing-dependent eligibility (STDE), which captures important experimental features in the striatum. The proposed computational model can recognize complex input patterns and consistently choose rewarded actions to respond to such sensorial inputs. Moreover, we assess the role different neuronal and network features, such as homeostatic mechanisms and lateral inhibitory connections, play in action-selection with the proposed model. The homeostatic mechanisms make learning more robust (in terms of suitable parameters) and facilitate recovery after rewarding policy swapping, while lateral inhibitory connections are important when multiple input patterns are associated with the same rewarded action. Finally, according to our simulations, the optimal delay between the action and the dopaminergic feedback is obtained around 300ms, as demonstrated in previous studies of RL and in biological studies.

Keywords: Striatum, Reinforcement learning, Spiking neural network, Dopamine, Eligibility trace, Spike-timing-dependent plasticity

1. Introduction

Animals learn to choose actions among many options by trial and error, thanks to the feedback provided by sparse and delayed rewards. Reinforcement learning (RL) serves as a theoretical framework for an agent, a system that acts based on received feedback, to learn to map situations to actions. This state-action mapping aims to maximize the performance of actions, mainly (but not exclusively) considering how rewarding or punishing the consequences of the actions are (Sutton et al., 1992). The basal ganglia (BG), a group of forebrain nuclei, are posited to play a critical role in

*Corresponding author
Email address: alvarogr@ugr.es (Álvaro González-Redondo)

action-selection based on RL (Grillner et al., 2005; Graybiel, 1998; Hikosaka et al., 2000; Gurney et al., 2001). However, the roles of recent findings, such as striatal spike-timing-dependent plasticity (STDP) models and striatal asymmetrical lateral connectivity, remain unclear. Investigating these interactions could improve our comprehension of the BG's role in RL, potentially leading to the development of more efficient bio-inspired reinforcement learning agents.

This study aims to explore the impact of homeostatic mechanisms and asymmetric lateral inhibitory connections on action-selection in the striatum. We use the RL framework to gain insights into the neural basis of decision-making and contribute to more biologically plausible basal ganglia models. Our model stands out from previous models in several ways: it does not require a critic or extra circuitry for a temporal difference signal, thereby simplifying the model and reducing computational complexity; additionally, it employs a spiking neural network with spike-time pattern representation that adapts well to varying pattern complexities in the pattern classification layer.

We propose a functional, biologically inspired striatum network model that incorporates dopamine-modulated spike-timing-dependent eligibility (STDE, Gurney et al. (2015)) and asymmetric lateral connectivity (Burke et al., 2017). This model improves upon existing striatum models by integrating homeostatic mechanisms, asymmetric lateral inhibitory connections, and the STDE learning rule, capturing essential experimental features found in the striatum.

In this article, we present a model that effectively processes complex input patterns in the context of reinforcement learning. We conduct multiple analyses to assess the interaction between the learning rule, homeostatic mechanisms, and lateral inhibitory connectivity patterns. By incorporating these elements, we strive to develop a comprehensive and biologically plausible striatum model that offers valuable insights. Our study examines the individual and combined effects of these factors, shedding light on the unique topology of the striatum network and its role in reinforcement learning tasks.

The main contributions and findings of this work are:

- A functional and biologically inspired network model of the striatum that integrates dopamine-modulated STDE, homeostatic mechanisms, and asymmetric lateral inhibitory connectivity, providing a more comprehensive and biologically plausible representation of the striatum's function.
- Analysis of the role of homeostatic mechanisms in making learning more robust and facilitating recovery after rewarding policy swapping.
- Investigation of the importance of lateral inhibitory connections when multiple input patterns are associated with the same rewarded action.
- The use of a spiking neural network with spike-time pattern representation that scales well with different pattern complexity, making the model suitable for a wide range of reinforcement learning tasks.
- Demonstration that the optimal delay between action and dopaminergic feedback occurs around 300 ms, which is consistent with previous reinforcement learning and biological studies.
- A model that does not require a critic, simplifying the learning process and reducing the need for additional circuitry.

1.1. Basal Ganglia Circuitry and Striatal Connectivity in Decision Making

The BG network is composed of several structures, grouped in inputs [being the striatum the best known, and populated by medium

spiny neurons (MSN)], intermediate layers [the external segment of the globus pallidus (GPe), and the substantia nigra pars compacta (SNc)] and output [substantia nigra pars reticulata (SNr)]. The information flows segregated through the BG circuits (DeLong et al., 1985; Parent and Hazrati, 1995). It has been proposed that the BG process a large number of cognitive streams or channels in parallel (Gurney et al., 2001), each of them representing a feasible action to be performed (Suryanarayana et al., 2019). According to recent research, this segregation through the entire cortico-BG-thalamic loop shows a very high specificity, down to almost neuron-to-neuron level (Hunnicutt et al., 2016; Foster et al., 2021). Thus, it seems feasible to impact behavior at different levels of detail. However, with the current biological evidence it is not exactly known how the activation of a channel maps to the corresponding behavior and we just assume here that these channels involve a decision making process.

The striatum, as the primary input of the basal ganglia, connects to the SNr via direct and indirect pathways, which are traditionally thought to promote and inhibit behavior, respectively. Each pathway crosses the striatum through different subpopulations of MSNs, expressing dopamine receptors D1 for the direct pathway and D2 for the indirect pathway. Recent genetic and optical studies on striatal circuits have allowed for testing classical ideas about the functioning of this system, but new models are needed to better understand the role of the striatum in learning and decision-making (Cox and Witten, 2019).

1.2. Spiking Neural Networks: Learning, Reward Modulation, and Striatal Connectivity

In recent decades, the use of biologically plausible computational models composed of spiking neurons able to learn a target function has demonstrated being increasingly successful (Taherkhani et al., 2020; Tavanaei et al., 2019). These models use discrete-time events (spikes) to compute and transmit information. As the specific timing of spikes carry relevant information in many biological contexts, these models are useful to understand how the brain computes at the neuronal description level. Combined with the use of local learning rules, these models can be implemented in highly efficient, low-power, neuromorphic hardware (Rajendran et al., 2019). Within this framework, learning from past experiences can be achieved using the STDP learning rule, a synaptic model featuring weight adaptation that has been observed in both biological systems (Levy and Steward, 1983) and the BG (Fino and Venance, 2010). The STDP also was demonstrated to be competitive in unsupervised learning of complex pattern recognition tasks (Masquelier et al., 2009; Garrido et al., 2016). The complexity of the patterns comes from their statistically equivalent activity level and from being immersed within a noisy stream of hundreds or thousands of inputs. These studies shown that an oscillatory stream of inputs reaching a population of spiking neurons enables a target post-synaptic neuron equipped with STDP to detect and recognize the presence of repetitive current patterns (Masquelier et al., 2009). The added oscillatory drive performs a current-to-phase conversion: the neurons that receive the most potent static current will fire the first during the oscillation cycle. This mechanism locks the phase of the spike time, facilitating the recognition of the previously presented patterns.

However, STDP-based learning systems tend to use statistical correlations to strengthen synaptic connections, resulting in the selection of the most frequent patterns at the expense of the most rewarding (Garrido et al., 2016). Thus, the STDP rule can be modified to drive the learning of patterns that statistically correlate with a reward signal (Izhikevich, 2007; Legenstein et al., 2008). In biological systems, unexpected rewards signal relevant stimuli dur-

ing learning by releasing dopamine (DA). More specifically, the reward signal is linked to the phasic modulation of dopaminergic neurons in the SNc and ventral tegmental area (Schultz, 2010), that sends reinforcement signals to the striatal neurons. These rewards do not need to happen instantly after the relevant stimulus; they can be delayed seconds, resulting in the distal reward and temporal credit assignment problems. In Izhikevich (2007); Legenstein et al. (2008), the authors suggest a reward-modulated STDP rule that enables a neuron to detect rewarded input patterns lasting milliseconds, even if the reward is delayed by seconds, by using the so-called eligibility trace. Also, based on the eligibility trace, Gurney et al. (2015) developed a synaptic learning rule called Spike-Timing-Dependent Eligibility (STDE) based on physiological data that captures many features found in the biological MSN of the basal ganglia. This model is more flexible than the previous STDP-like rules as different learning kernels can be used depending on the amount and type (reward or punishment) of reinforcement received. Although the authors did not include some important BG features like the GPe nucleus or a corticostriatal loop, their model successfully learned to select an action channel driven by stronger cortical input, based only on the timing of the input and the reward signal.

Another relevant feature of the striatum is its connectivity. Burke et al. (2017) proposed a model of asymmetric lateral connectivity in the striatum that tries to explain how different clusters of striatal neurons interact and which role they play in information processing. This model accounts for the *in vivo* phenomenon of co-activation of sub-populations of D1 or D2 MSNs, which seems paradoxical as each subpopulation projects to behaviorally opposite pathways (direct and indirect, respectively). This structured connectivity pattern is determined by lateral inhibition between neurons that belong to the same channel and between neurons within different channels but accounting for the same receptor type (D1 or D2). The authors also include asymmetrical connections with more intensive intra-channel inhibition from D2 to D1 neurons than in the opposite direction. This pattern resulted in synchronized phase-dependent activation between MSN D1 and D2 neuron groups that belong to different channels.

1.3. Contribution

All the previous ideas are important pieces of the process of goal-oriented learning but further research is required as their respective roles and how they complement each other are still not well understood. The combination of the STDE rule within a network with asymmetrically structured lateral inhibition has not been studied before, and some relevant conclusions emerge from this specific study. In this article, we design and study a functional and biologically inspired model of the striatum. Our approach is based on spike time representation of complex input patterns and integrates dopamine modulated STDE and asymmetric lateral connectivity, among other mechanisms. This model learns to select the most rewarding action to complex input stimuli through RL. The proposed model has been demonstrated to be capable of recognizing input patterns relevant for the task and consistently choosing rewarded actions in response to that input. We performed numerous analyses to measure and better understand the interaction between the learning rule with homeostatic mechanisms and the lateral inhibitory connectivity patterns. By measuring the single and combined effects of these factors in the learning process, we want to shed light on how the particular topology of the striatum network facilitates the resolution of RL tasks.

2. Methods

Aiming to implement a RL framework in a biologically plausible striatum model, we started designing a task where the agent has to learn how

to map different input patterns into actions based on the reward signal delivered by the environment. We implemented a network model of the striatum capable of learning this task. This system behaves like a RL agent and can solve action-selection tasks.

The methods section is structured as follows: we first define the neuron and synapse models, input pattern generation, and networks structures used in our experiments. Then we describe the experimental design used with the network model and how we measure its learning capability. In Supplementary Materials we also explain both a previous experiment and a simpler model we made to test the viability of the combination of oscillatory inputs, STDE and homeostatic rules that we employed in the final network model.

2.1. Computational models

2.1.1. Neuron models

We used conductance-based versions of the Leaky-Integrate and Fire (LIF) neuron model (Gerstner and Kistler, 2002) as it is computationally efficient and captures certain biological plausibility. We use this model in every layer of the network, but with different parameters. We classify the neuron types according to the layer they belong to: cortical neurons for the input, striatal neurons (divided in two subpopulations according to which DA receptor express, D1 or D2) for the learning layer, and action neurons for the output. There is also a dopaminergic neuron that receives the rewards and punishments. The parameters used for each type were manually tuned to obtain reasonable firing rates. For the cortical neurons we used a number of spikes per input cycle (with 8 cycles per second) close to Masquelier et al. (2009) and Garrido et al. (2016) (see details about the input protocol in section 2.1.2). For the striatal neurons, we tuned the parameters to obtain a mean firing rate of around one spike per second to be within biological ranges (Miller et al., 2008) but with activity peaks of two or three spikes per input cycle (16-24 spikes per second). The action neurons (an integrative population that outputs the agent's behavior) are tuned to fire every input cycle if they receive enough stimulation from its channel (at least two more spikes from D1 neurons than D2 neurons each cycle). The dopamine neuron was tuned to have a firing range from 50 to 350 spikes per second, with these unrealistic values chosen for performance (instead of simulating a bigger dopaminergic population). The parameters used for each neuron type are shown at supplementary table 1.

2.1.2. Input and oscillatory drive

In the input generation procedure (Masquelier et al., 2009; Garrido et al., 2016) we consider a trial as a segment of time of the simulation where we present some input stimuli to the network. The length of each trial is taken from a uniform random distribution between 100 and 500ms. An input stimulus represents a combination of 2000 input current values conveyed one-to-one to a set of cortical neurons of the same size (Fig. 8A). An input *pattern* is a combination of current values which target precisely the same cortical neurons every time the input pattern is presented for the entire simulation. For every time bin, one or no pattern is presented. Only half of the cortical neurons (1000) are pattern-specific when presenting a specific pattern, while the other half receives random current values. The cortical neurons specific for each pattern are selected at the initialization. When no pattern is presented, all the cortical neurons receive random current values. Two thousand current-based LIF cortical neurons transform the input current levels into spike activity. These neurons have a firing rate between 8 to 40 spikes per second due to the sum of the input current values (ranged from 87% to 110% of the cortical neuron rheobase currents) and an oscillatory drive at 8Hz feeding these neurons (with an amplitude of 15% of the rheobase current of the cortical neurons). This oscillatory drive turns the input encoding from analogical signal to phase-of-firing coding (Masquelier et al., 2009) by locking the phase of the cortical spikes within the oscillatory drive, as shown in Fig. 8B. By using these parameters, the cortical neurons fire between 1 and 5 spikes per cycle.

2.1.3. Spike-Timing-Dependent Eligibility (STDE) learning rule

We implemented a version of the STDE learning rule (Gurney et al. (2015)), a phenomenological model of synaptic plasticity. This rule is similar to STDP, but the kernel constants are DA-dependant (that is, different values are defined for low DA and high DA values, and interpolated for DA values inbetween, as shown in Fig. 1 and Supplementary Fig. 9Ai and Aii). STDE is derived from *in vitro* data and predicts changes in direct and indirect pathways during the learning and extinction of sin-

gle actions. Throughout, we used the following parameters and procedures unless we specified otherwise. The kernel shape is defined by the parameters k_{DA}^{SPK} with $SPK \in \{+, -\}$ being the spike order pre-post for applying k_{DA}^{+} and post-pre for applying k_{DA}^{-}, respectively, and $DA \in \{hi, lo\}$ being the high- or low-DA cases, resulting in four parameters in total: k_{hi}^{+}, k_{lo}^{+}, k_{hi}^{-} and k_{lo}^{-}. We obtained these learning kernel constant values by hand-tuning for both MSN D1 and D2 cases (see Supplementary Fig. 9 and supplementary table 2). As in the classic STDP learning rule, the weight variation in STDE is calculated for every pair of pre- and post-synaptic spikes and decays exponentially with the time difference between the spikes (Fig. 1). We use time constants $\tau = 32$ ms and the weights values are clipped to $[0, 0.075]$.

Our implementation of STDE uses elegibility traces that decay exponentially to store the potential weight changes, similarly to Izhikevich (2007). Following Gurney et al. (2015) we have two different eligibility traces per synapse, c^+ and c^- for spike pairs with positive and negative timing respectively, updated for every pair of pre- and post-synaptic spikes at times t_j and t_i as in equations (1) and (2):

$$\delta c^+ = \left(\alpha k_{hi}^+ + \overline{\alpha} k_{lo}^+\right) \cdot e^{\frac{t_j - t_i}{\tau_{eli}}} \text{ if } t_j \leq t_i \qquad (1)$$

$$\delta c^- = \left(\alpha k_{hi}^- + \overline{\alpha} k_{lo}^-\right) \cdot e^{-\frac{t_j - t_i}{\tau_{eli}}} \text{ if } t_j > t_i \qquad (2)$$

with $\overline{\alpha} = 1 - \alpha$, α been a value dependent of DA that we define in equation 3, and τ_{eli} been the eligibility trace time constant with a value twice the length of the mean reward delay. Overall plastic change at a single synapse is then the sum of contributions from both c^+ and c^-, scaled by a learning rate factor $\eta = 0.002$.

The level of DA in the system is determined by one neuron that fires at high (and unrealistic) rates for computational simplicity, representing a population of neurons from the SNc. This neuron fires spontaneously at a baseline frequency of 200Hz. The environment (i.e., the application of rewarding policies during the experiment) injects positive (or negative) current in the dopaminergic neuron when rewards (or punishments) are applied to the model, resulting in the firing rate of this neuron ranging between 50Hz and 350Hz. All plastic synapses share a global DA level d that decays exponentially with temporal constant $\tau_{da} = 20$ms. For each spike emit-

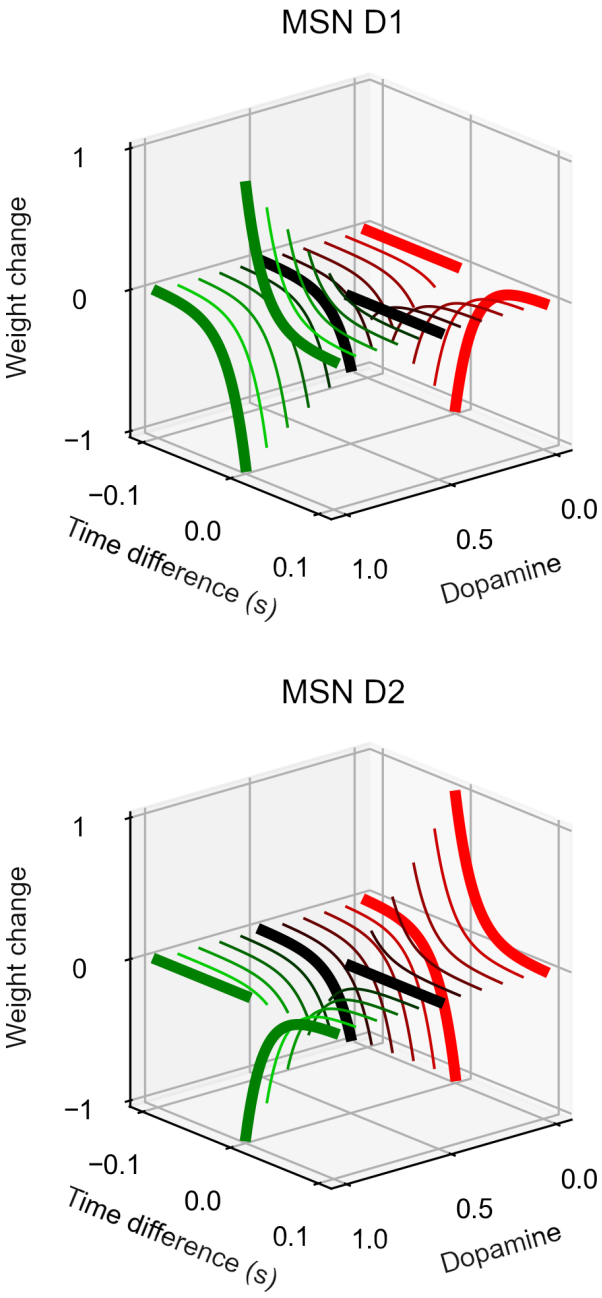

Figure 1: Kernels used for STDE synapses of MSN D1 (top) and D2 (bottom), showing the weight change depending on the time difference between pre- and post-synaptic spikes and dopamine. Thick lines represent kernels at dopamine minimum, normal, and maximum values (red, black, and green, respectively). Thin lines are interpolations of these values.

ted by the dopaminergic neuron, d is increased by $\frac{1}{\tau_{da}}$ with 200-ms delay.

Our implementation of STDE uses the linear mixing function α in equation (3), clipped to $[0, 1]$, to smoothly morph between kernels with low and high DA:

$$\alpha = \frac{d - d_{min}}{d_{max} - d_{min}} \tag{3}$$

where d_{min} and d_{max} are the minimum and maximum values of DA considered. We use this equation for computational efficiency instead of the Naka-Rushton function used in Gurney et al. (2015) (the authors also noted that this is not a requirement, as long as the mixing function was increasingly monotonic and saturating). The function is bounded to values of DA firing rate between 50 and 350Hz, with the baseline at 200Hz.

2.1.4. Homeostatic mechanisms

During learning, in some cases, the neurons can stop firing indefinitely due to a learning history leading to the wrong parameters. Neuron activity can also die by sudden changes in the reward policy, leaving the state of the synaptic weights ill (not representing any stimuli and not getting enough input to fire by chance). To recover neurons from this state, we added two different homeostatic mechanisms, one at the synaptic level and one at the neuron level. Although one or the other is enough to avoid the ill-states, we saw in our tests that we recovered faster and more reliably by using both.

The synapses implementing the STDE included a non-Hebbian strengthening in response to every pre-synaptic spike. For each arriving spike, the synaptic weight increases by $C_{pre} = \eta \cdot 4 \cdot 10^{-4}$. This non-Hebbian strengthening is added to enable the recovery of low-bounded synapses (e.g., after a rewarding policy switch). Although the rewarding policy does not change in the network experiment, this homeostatic mechanism also benefits the complete network model learning (more details in section 5.2.2 and Supplementary Fig. 14).

In order to avoid neurons to become permanently silent during learning, we include adaptive threshold to our neuron models based on Galindo et al. (2020) according to the following equation:

$$\frac{dV_{th}}{dt} = -\frac{V_{th} - E_{leak}}{\tau_{th}} \tag{4}$$

where V_{th} represents the firing threshold at the current time, E_{leak} is the resting potential of the neuron, and τ_{th} is the adaptive threshold time constant. According to equation 4, in the absence of action potentials, the threshold progressively decreases towards the resting potential, facilitating neuron firing. When the neuron spikes, the firing threshold increases a fixed step proportional to the constant C_{th} as indicated in equation 5, making neuron firing more sparse.

$$\delta V_{th} = \frac{C_{th}}{\tau_{th}} \tag{5}$$

2.1.5. Striatum network model

The network model of the striatum (Fig. 3A) contains two channels (channel A and channel B, each one representing a possible action). Every channel contains two same-sized subpopulations (D1 and D2 neurons, respectively) of striatal-like neurons (in total, 16 neurons per channel) and one so-called *action neuron* that integrates excitatory activity from D1 neurons and inhibitory activity from D2 neurons. This design simplifies the biological substrate in which all MSN are inhibitory, but we implemented the network computation by considering the net effect of each neuron type on behavior. Biological MSN D1 neurons inhibit SNr, which promotes behavior, and MSN D2 neurons inhibit GPe, which, in turn, inhibit SNr with the total effect of decreasing behavior (Fig. 3A).

Our striatum model implements lateral inhibition within each MSN D1 population, within each MSN D2 population, between MSN D1 and MSN D2 populations within the same channel, and between the MSN populations associated with different action channels. Inspired by Burke et al. (2017), we used an asymmetrical structured pattern of connectivity (Fig. 5E in (Burke et al., 2017), and adapted here in Fig. 2). Following this connectivity pattern, we added lateral inhibition between neurons that belong to the same channel and between those that belong to different channels but use the same dopaminergic receptor D1 or D2 (with stronger inhibition from D2 to D1 neurons than in the opposite direction). Since the small size of the network under study and the small weight of the D1 to D2 MSN connections, the overall contribution of these connections was neglectable, so we decided not to include them in our simulations as we see no significant impact on previous simulations.

The environment generates the reinforcement signal based on comparing the chosen and the expected action and then delivers it to the dopamin-

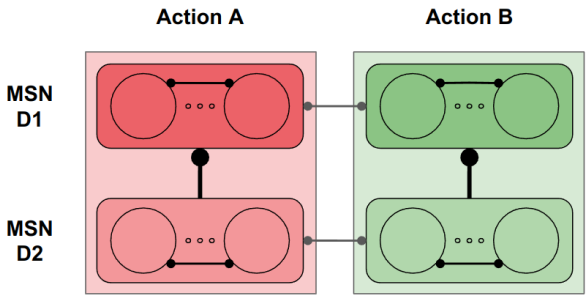

Figure 2: Connectivity pattern used for the lateral inhibition, inspired on Burke et al. (2017). Two channels (action A and action B) are shown, each with two populations of D1 and D2 MSN.

ergic neuron. Rewards are excitatory, and punishments are inhibitory inputs to this neuron. The dopaminergic modulatory signal is global and delivered to every STDE connection from cortical layer to striatal neurons (Fig. 3A). It is important to note that this model does not implement a critic (commonly used in actor-critic frameworks (Sutton et al., 1992)), so there is no reward prediction error signal.

2.2. Experimental design

We first validated the proposed learning mechanisms with a simpler network model of only one neuron and a easier experimental task, as can be seen in Supplementary Methods 5.1 and Supplementary Results 5.2.

The action-selection task used to test the model (Fig. 3B) works as follows: the agent has two possible actions to choose, A or B. An action is selected if the activity balance of its D1 and D2 neurons is biased to D1 in two spikes at least in one cycle (making the corresponding action neuron spike). The agent can do none, both, or any of them at a time. The input stream contains five different non-overlapping input patterns, each one presented 16% of the time (80% in total). The policy used to give rewards (excitation) and punishments (inhibition) to the agent (dopaminergic neuron) is the following. When pattern 1 or 2 is present, the agent is rewarded if action A is selected (action A neuron fires during the pattern presentation and action B neuron does not fire) but punished if action B is selected. When pattern 3 or 4 is present, the agent is rewarded if action B is selected but punished if action A is selected. When pattern 5 is present, the agent is punished if it selects action A or B. This policy applies no punishment or reward to the agent during noisy inputs, whatever the action taken is. In case of spiking both action neurons during a reinforced input, the network is punished.

2.3. Performance measurement

In the action-selection task we measure the performance of the models by calculating the percentage of correct action choices (i.e. the learning accuracy). This measure is widely used in classification problems when the objective is to describe the accuracy of a final map process (Stehman, 1997). To do so, for each pattern presentation we store the rewarded (expected) action in response to the presented pattern, and the finally selected (chosen) one during that pattern presentation. We only consider in the calculation those trials in which some reward or punishment can be delivered, ignoring those intervals with no repeating patterns conveyed to the inputs (only noisy inputs). We consider that an action has been taken if the corresponding action neuron has spiked at least once during the pattern presentation. Conversely, we consider that no action has been taken if none of the action neurons spikes during the same duration. In order to obtain an estimation of the temporal evolution of the accuracy we use a rolling mean of the last 100 values.

3. Results and discussion

We did extensive testing of the learning mechanisms we proposed. Some of these results demonstrate that the combination of STDE learning rule and homeostatic mechanisms allow learning (and re-learning) of rewarded patterns, or that there is no effect of the reward delay and the frequency of the input pattern on the learning process, among others. However, as they are not the main concern for this article, they are placed in the Supplementary Results 5.2 section for further examination.

The main results and discussion are structured as follows: we first show the general behavior of the network. Then we study the effect

of the lateral connectivity pattern on the performance and the way neurons are processing information. Finally, we put our results in context by comparing our model with previously proposed models in the literature.

3.1. General network behavior

During the simulation of the action-selection task, each action group neuron becomes overall active in response to the presentation of the associated patterns as shown in the raster plots (Fig. 3C and D) and the activity balance for the action neuron groups (Fig. 3E), producing mainly dopaminergic rewarding (Fig. 3F). The action accuracy reveals steady-state performance after 200 seconds of simulations (Fig. 3G). According to these results, our combination of STDE learning rule (Gurney et al., 2015) with homeostatic mechanisms and an oscillatory input signal in a cortico-striatal model learns to accurately select the most rewarding action.

The way our network learns to associate the corresponding input stimulus with subpopulations of D1 and D2 neurons in channel A or channel B is the following: If the agent takes the right action for a specific input pattern, the environment delivers a reward with some delay (high DA level in Fig. 3F). This reward potentiates the synapses between the cortical layer and the action-associated D1 sub-population, resulting in more frequent firing. On the other hand, if the agent takes a wrong action, then it receives a punishment sometime later (low DA level in Fig. 3F). This punishment weakens the synapses from the cortical layer to the action-associated D1 sub-population while strengthening the corresponding synapses to the D2 (inhibitory) sub-population of the same channel. This learning process makes the agent stick to the rewarded action and switch to a different one when punished. For the specific case when the environment punishes any action during a stimulus presentation, both D2 sub-populations increase their activity, and both action neurons remain silent.

The proposed model shows how combining two complementary dopamine-based STDE learning rules (Fig. 1) can facilitate the association between sensorial cortical inputs and rewarded actions with arbitrary rewarding policies. Previously, the STDE rule had been shown to be capable of learning to select an action channel driven by stronger cortical input (Gurney et al., 2015), and here we show that this rule can also be used to learn inputs defined by the specific timing of their spikes (as all the inputs have the same average firing rate). This represents a higher complexity task and illustrates how STDE can be efficiently used for spike time pattern representation.

The model also is completely bioplausible, as all the mechanisms used have been described in biological systems: DA induces bidirectional, timing-dependent plasticity at MSNs glutamatergic synapses (Shen et al., 2008), *in vitro* pyramidal neural recordings are consistent with simulations of adaptive spike threshold neurons, and they lead to better stimulus discrimination than would be achieved otherwise (Huang et al., 2016), and rat hippocampal pyramidal neurons *in vitro* can use rate-to-phase transform (McLelland and Paulsen, 2009). Detailed discussion on the role of the homeostatic mechanisms can be found in Supplementary Materials.

3.2. Effect of lateral inhibition patterns and task complexity

Once we have demonstrated how the striatal network can support RL, we wondered to what extent the connectivity pattern of the lateral inhibition in the striatum could impact the learning capabilities. So that we extensively explored different versions of connectivity.

We first study if there is any relationship between the connectivity pattern and difficulty of the task. We organized the lateral inhibitory connections in two groups: intra-channel (inhibitory connections from D2 MSNs to D1 MSNs within the same channel) and

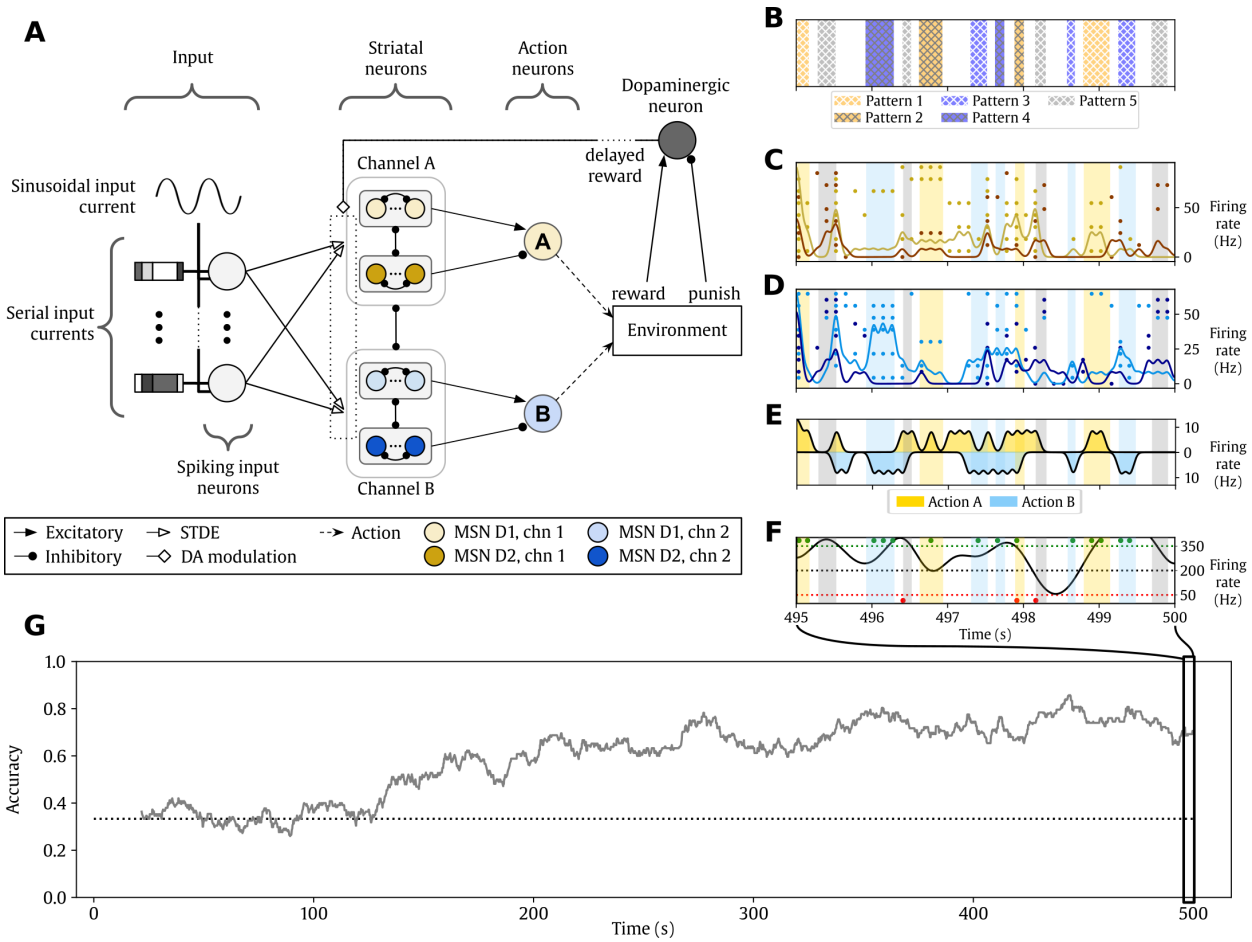

Figure 3: Cortico-striatal network solving a RL task. **A**. Structure of the network. See section 2.1.5 for a detailed explanation. **B-F**. The activity of the network during the last 5 seconds of simulation. Background color indicates the reward policy (yellowish colors, action A is rewarded and B is punished; bluish colors, action B is rewarded and A is punished; grey, any action is punished). **B**. Input pattern conveyed to the cortical layer. **C**. Raster plot of the channel-A action neurons. Yellow dots represent MSN D1 spikes, and orange dots are MSN D2 spikes. **D**. Raster plot of channel B. Cyan dots represent MSN D1 spikes, and dark blue dots are MSN D2 spikes. **E**. Action neuron firing rates. The middle horizontal line represents 0 Hz. Action A and B activity are represented in opposites directions for clarity. Action A neuronal activity increases in yellow zones while action B neuronal activity in cyan intervals. **F**. Firing rate of the dopaminergic neuron (black line). Dotted horizontal lines indicate the range of DA activity considered: black is the baseline, green is the maximum reward, and red represents the maximum punishment. Dots indicate rewards (green) and punishment (red) events delivered to the agent. **G**. Evolution of the learning accuracy of the agent, see section 2.3 for further details. The dotted line marks the accuracy level by chance.

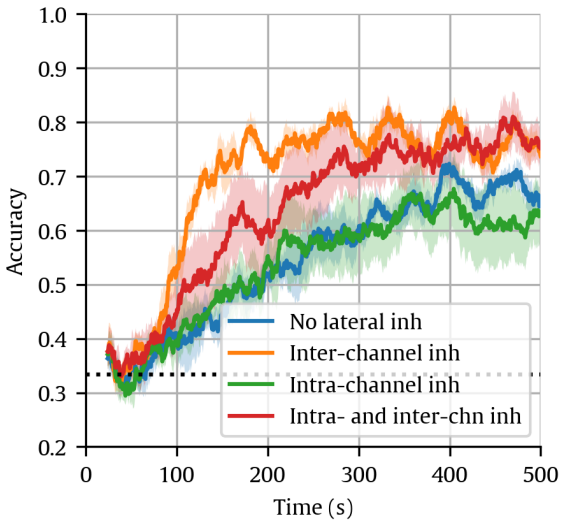

Figure 4: Effect of the lateral inhibitory connectivity on the performance during a simpler version of the RL task. The horizontal dotted line represents the accuracy obtained by a random agent. The curves represent the mean and the standard error of the mean of the evolution of each agent during the task (n=5).

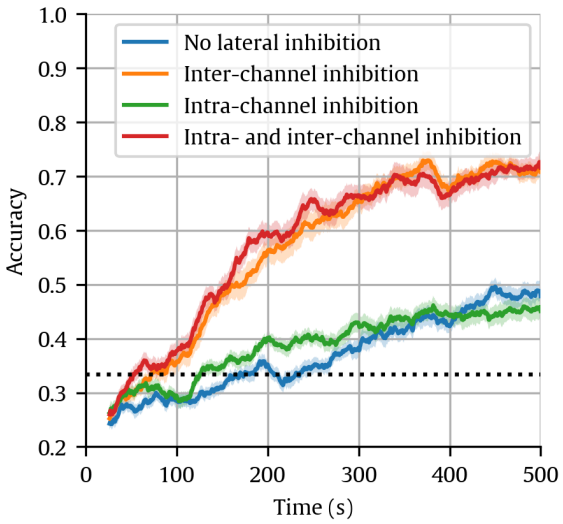

Figure 5: Effect of the lateral inhibitory connectivity on the performance during the normal RL task. The curves represent the mean accuracy and the shaded areas represent the standard error (n=30). Four different configurations are tested, depending on the presence of two types of lateral connectivity: intra- and inter-channel inhibition. The horizontal dotted line represents the accuracy obtained by a random agent with no learning mechanisms.

inter-channel (inhibitory connections between D1 MSNs of different channels, and between D2 MSNs of different channels). We obtained four possible subsets of connectivity patterns by keeping or removing each connection type (Fig. 2). We used three difficulty levels for the task: easy, normal and hard. The easy task uses only one stimulus associated with each action (stimulus 1 to action A, stimulus 2 to action B, stimulus 3 to no action). The normal task uses two stimulus per action, and one no-go stimulus. The hard task uses four stimuli per action, and two no-go stimuli.

The results of the easy version of the experiments are shown in the Fig. 4. The models without inter-channel inhibition work worse, as they stabilize with lower values of accuracy. The models with inter-channel inhibition seem to reach a similar level of accuracy but the intra-channel inhibition seems to reduce the learning rate.

In the normal version of the task, we again obtained the best learning performance when using the inter-channel lateral inhibition with asymmetrical structured connection pattern, and the difference increased. In this case, there is no apparent effect in of the intra-channel lateral inhibition in this task (Fig. 5). According to our simulations, lateral inter-channel inhibition facilitates the emergence of one action-related channel over the other one in a winner-take-all manner, as expected.

We saw in previous experiments that the inter-channel lateral inhibition is always increases accuracy, so we will use it always in the following tests. In the hard task we obtained small but significant differences: The accuracy of the network improves faster with the intra-channel lateral inhibition (see Fig. 6). Also, apparently the network with the intra-channel inhibition settled in a more stable regime as it maintains its performance, compared with the network without this intra-channel inhibition which slowly degrades (Supplementary

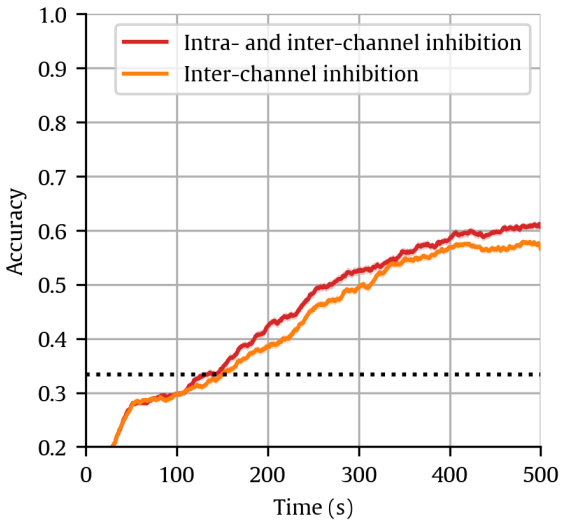

Figure 6: Effect of the intra-channel lateral inhibitory connectivity on the performance during a harder version of the RL task. The horizontal dotted line represents the accuracy obtained by a random agent. The curves and the filling color represent the mean, the standard error of the mean, respectively, of the evolution of each agent during the task (n=150), simulated for 500 seconds.

Fig. 13). The results so far suggest that both connectivity patterns contribute to a reliable action-selection paradigm.

Taking these results together, it seems that when we use several stimuli associated with each action, intra-channel inhibition improves the RL action selection task. However, when only one stimulus is associated with each action, this intra-channel inhibition does not impact learning performance. When compared with the results in Fig. 5 and 6, it seems that the intra-channel lateral inhibition improves the learning capabilities only with a harder task, but when the task is too simple then the intra-channel connection increases the learning time.

We also explored the effect of connectivity patterns of lateral inhibition different from the proposed by Burke et al. (2017)), by adding or removing lateral connections within a channel, within each subpopulation, and between subpopulations of the same channel. All variations from the original resulted in reduced learning performance (Supplementary Fig. 15). In this Figure, the curve #5 represents the network with both lateral inhibition in D1 layer and D2 layer, as well as intra- and inter-channel lateral inhibition. This structure (similar to the one proposed by Burke et al. (2017)) obtains the best accuracy.

3.3. Effect of intra-channel lateral inhibition on neuronal specialization

Intra-channel inhibition seems to facilitate learning in more complex tasks, possibly because it enhances neuron specialization. We saw a strong reduction of correlation at time difference $\delta t = 0$ between action A and B D1 sub-populations caused by intra-channel inhibition (data not shown), but this does not seem to justify the improved accuracy for more complex tasks.

Then, we hypothesized that intra-channel inhibition could encourage neuron specialization to specific cortical patterns. We tested this idea by analyzing the preferred stimuli for each neuron after the learning process (Fig. 7), and obtained the opposite result: the intra-channel lateral inhibition affects D1 neurons by forcing them to share more evenly their activity over several stimuli, in addition to reducing their average activity. This is in contrast with the network without intra-channel lateral inhibition, where the activity is more focused on the favorite stimuli and has higher mean activity.

According to these results, although individual neurons of the network with intra-channel inhibition have less precise representation of individual sensorial stimuli, these models have higher precision to associate rewarding actions. This can be explained assuming some sparse representation of the stimuli, where the simultaneous firing of several (but not many) neurons are needed to indicate the presence of an input stimuli. This more sparse representation emerges due to the combination of stronger inhibition and the homeostatic mechanisms: a neuron avoids firing when it is inhibited, so

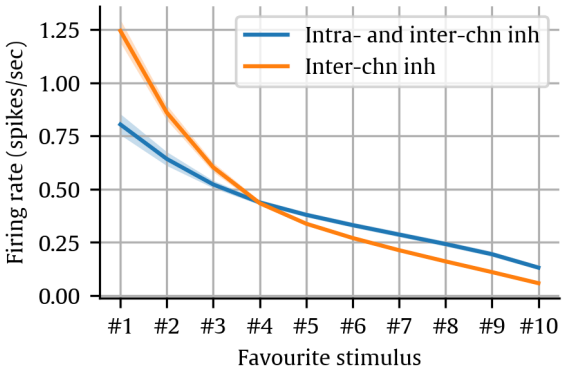

Figure 7: Effect of the intra-channel lateral inhibitory connectivity on the firing rate pattern on their preferred stimuli of D1 neurons. Higher and more specialized firing patterns occur in networks without intra-channel lateral inhibition, while more sparse representations occur in networks with it. Lines and shaded areas represent mean and 95% confidence intervals of the mean ($n = 150$), respectively.

the homeostatic mechanisms tend to compensate for this activity reduction by increasing its chances to fire in response to several stimuli. This sparse representation has been suggested to facilitate sensorial pattern recognition in other brain areas, such as the cerebellar cortex, the mushroom body, and the dentate gyrus of the hippocampus (Cayco-Gajic and Silver, 2019).

In the context of our model, the sparse representation due to intra-channel inhibition plays a role in the action selection process, which can be seen as a form of classification. Here, the goal is not to classify stimuli per se, but to assign stimuli to appropriate actions. The sparse coding helps to achieve more efficient and robust action selection by reducing the overlapping between representations of different sensorial states, minimizing interference, and enabling more reliable decision-making.

3.4. Comparison with previous models of reinforcement learning and basal ganglia

We presented a point-neuron model of the BG that can solve complex action-selection tasks using a RL paradigm. We do so by using multiple mechanisms proposed in the literature: the STDE learning rule that implements synaptic modification in cortex-MSN connections (Gurney et al., 2015), combined with homeostatic mechanisms (Galindo et al., 2020) and an oscillatory input signal (Masquelier et al., 2009; Garrido et al., 2016) in a network with asymmetrical structured lateral inhibition (Burke et al., 2017) can rapidly and consistently learn to detect the presence of rewarded input patterns. These processes have been described in biological systems and here proved to be robust.

Simpler STDP-like rules have been used for RL tasks (Izhikevich, 2007; Legenstein et al., 2008), but they were employed in simpler networks, single neurons, and simple tasks. Beyond the state-action mapping role proposed in this article for the striatum, other theories exist about the action decision process. However, computational models of BG in the literature have considerably evolved during the last two decades (Rubin et al., 2021), and there is still no consensus about how to achieve goal-oriented learning in a BG model. Previous models ranged from those with action-selection features but no learning (Beiser et al., 1997; Gillies and Arbuthnott, 2000; Humphries et al., 2006; Lo and Wang, 2006a; Berns and Sejnowski, 1998; Gurney et al., 2001; Sen-Bhattacharya et al., 2018; Frank, 2006; Ratcliff and Frank, 2012; Bogacz, 2007) (but see (Frank, 2005)) to simple forms of learning, with RL (Bogacz and Larsen, 2011), rate-based learning rules (Hong and Hikosaka, 2011), or based on modulated STDP with eligibility traces (Humphries et al., 2009; Gurney et al., 2015; Baladron et al., 2019). These models considered direct and indirect pathways (as "selection" and "control" routes, respectively), composed of MSN D1 and D2 striatal neurons controlling GPe and SNr. Many models assume that the BG work as an actor-critic model (Bogacz and Larsen, 2011; O'Doherty et al., 2004), and actor-critic frameworks have

been used for RL tasks like maze navigation (Frémaux et al., 2013; Potjans et al., 2009; Vasilaki et al., 2009) and cartpole (Frémaux et al., 2013). More biologically-constrained models of the BG have been proposed to explain the origin of diseases like Parkinson's disease (Lindahl and Kotaleski, 2016) and the role of specific interneurons (Goenner et al., 2021) or pathways (Girard et al., 2021) during action-selection. Recent accumulation-to-bound models describe the decision process as an accumulation of evidence for each alternative action until a decision threshold is exceeded in one of these actions (Mulder, 2014). It would be interesting to explore how these models could be incorporated with the proposed model, potentially requiring additional brain areas. In this regard, some models incorporate recurrent activity loops with the cortex through the thalamus (Lo and Wang, 2006b).

Moreover, we acknowledge that similar models can already deal with more complex action-selection tasks than the one used in this work, such as cart-pole, inverted pendulum, or simple mazes (Frémaux et al., 2013). However, there exist some important differences between their model and the one proposed in this article. First, our network does not include a critic. Second, their learning rule requires a temporal difference (TD) signal that would need additional circuitry. Third, their model requires an additional place-cell layer with unsupervised learning to represent complex input patterns. However, it remains as a future work to embed the network model into a closed-loop experimental setup requiring continuously graded output (instead of selecting an action in a discrete set of possibilities). This way, the model could deal with a larger set of RL tasks. In our case, we have integrated a spiking neural network with spike-time pattern representation that scales well with different patterns complexity at the pattern classification layer. Future work will explore how our model could be extended for such complex action control frameworks.

4. Conclusion

In this article we tested the respective roles in learning of the different mechanisms used during our simulations: homeostatic mechanisms make the neurons change their response to compensate for long-lasting changes in the input level, making learning faster and more robust to the configuration. The asymmetrical lateral inhibition consistently outperformed other connectivity configurations. By adding intra-channel lateral inhibition to the network model, we induced the channels to generate a sparse representation of each stimulus relevant for the task. This made the network less prone to errors as the model had to recruit more neurons to take an action. Lastly, by segregating striatal and action neurons in independent channels for each action and incorporating MSN D1 (Go neurons) and MSN D2 (No-Go) sub-populations with different learning kernels, the model effectively learned arbitrary mappings from sensorial input states to action output in a two-choice action-selection task. MSN D1 neurons and MSN D2 neurons cooperatively facilitated action selection with contrary effects; MSN D1 neurons learned to potentiate preferred actions while MSN D2 neurons learned to inhibit non-preferred actions.

Acknowledgements

This research is supported by the Spanish Grant INTSENSO (MICINN-FEDER-PID2019-109991GB-I00), Regional grants Junta Andalucía-FEDER (CEREBIO P18-FR-2378 and A-TIC-276-UGR18). This research has also received funding from the EU Horizon 2020 Framework Program under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3) and the EU Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 891774 (NEUSEQBOT). Additionally,

the main author has been funded with a national research training grant (FPU17/04432). Finally, this research was also supported by the Vetenskapsrådet (VR-M-2017-02806, VR-M-2020-01652); Swedish e-science Research Center (SeRC); KTH Digital Futures.

References

Baladron J, Nambu A, Hamker FH. The subthalamic nucleus-external globus pallidus loop biases exploratory decisions towards known alternatives: a neuro-computational study. European Journal of Neuroscience 2019;49(6):754–67.

Beiser DG, Hua SE, Houk JC. Network models of the basal ganglia. Current opinion in neurobiology 1997;7(2):185–90.

Berns GS, Sejnowski TJ. A computational model of how the basal ganglia produce sequences. Journal of cognitive neuroscience 1998;10(1):108–21.

Bogacz R. Optimal decision-making theories: linking neurobiology with behaviour. Trends in cognitive sciences 2007;11(3):118–25.

Bogacz R, Larsen T. Integration of reinforcement learning and optimal decision-making theories of the basal ganglia. Neural computation 2011;23(4):817–51.

Burke DA, Rotstein HG, Alvarez VA. Striatal local circuitry: a new framework for lateral inhibition. Neuron 2017;96(2):267–84.

Cayco-Gajic NA, Silver RA. Re-evaluating circuit mechanisms underlying pattern separation. Neuron 2019;101(4):584–602.

Cox J, Witten IB. Striatal circuits for reward learning and decision-making. Nature Reviews Neuroscience 2019;20(8):482–94. doi:10.1038/s41583-019-0189-2.

DeLong MR, Crutcher MD, Georgopoulos AP. Primate globus pallidus and subthalamic nucleus: functional organization. Journal of neurophysiology 1985;53(2):530–43.

Fino E, Venance L. Spike-timing dependent plasticity in the striatum. Frontiers in synaptic neuroscience 2010;2:6.

Foster NN, Barry J, Korobkova L, Garcia L, Gao L, Becerra M, Sherafat Y, Peng B, Li X, Choi JH, Gou L, Zingg B, Azam S, Lo D, Khanjani N, Zhang B, Stanis J, Bowman I, Cotter K, Cao C, Yamashita S, Tugangui A, Li A, Jiang T, Jia X, Feng Z, Aquino S, Mun HS, Zhu M, Santarelli A, Benavidez NL, Song M, Dan G, Fayzullina M, Ustrell S, Boesen T, Johnson DL, Xu H, Bienkowski MS, Yang XW, Gong H, Levine MS, Wickersham I, Luo Q, Hahn JD, Lim BK, Zhang LI, Cepeda C, Hintiryan H, Dong HW. The mouse cortico–basal ganglia–thalamic network. Nature 2021;598(78797879):188–194. doi:10.1038/s41586-021-03993-3.

Frank MJ. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism. Journal of cognitive neuroscience 2005;17(1):51–72.

Frank MJ. Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. Neural networks 2006;19(8):1120–36.

Frémaux N, Sprekeler H, Gerstner W. Reinforcement learning using a continuous time actor-critic framework with spiking neurons. PLoS computational biology 2013;9(4):e1003024.

Galindo SE, Toharia P, Robles ÓD, Ros E, Pastor L, Garrido JA. Simulation, visualization and analysis tools for pattern recognition assessment with spiking neuronal networks. Neurocomputing 2020;400:309–21. doi:10.1016/j.neucom.2020.02.114. arXiv:2003.06343.

Garrido JA, Luque NR, Tolu S, D'Angelo E. Oscillation-Driven Spike-Timing Dependent Plasticity Allows Multiple Overlapping Pattern Recognition in Inhibitory Interneuron Networks. International Journal of Neural Systems 2016;26(05):1650020. doi:10.1142/S0129065716500209.

Gerstner W, Kistler WM. Spiking neuron models: Single neurons, populations, plasticity. Cambridge university press, 2002.

Gillies A, Arbuthnott G. Computational models of the basal ganglia. Movement Disorders 2000;15(5):762–70.

Girard B, Lienard J, Gutierrez CE, Delord B, Doya K. A biologically constrained spiking neural network model of the primate basal ganglia with overlapping pathways exhibits action selection. European Journal of Neuroscience 2021;53(7):2254–77.

Goenner L, Maith O, Koulouri I, Baladron J, Hamker FH. A spiking model of basal ganglia dynamics in stopping behavior supported by arkypallidal neurons. European Journal of Neuroscience 2021;53(7):2296–321.

Graybiel AM. The basal ganglia and chunking of action repertoires. Neurobiology of Learning and Memory 1998;70(1):119–36. doi:https://doi.org/10.1006/nlme.1998.3843.

Grillner S, Hellgren J, Ménard A, Saitoh K, Wikström MA. Mechanisms for selection of basic motor programs – roles for the striatum and pallidum. Trends in Neurosciences 2005;28(7):364–70. URL: https://www.sciencedirect.com/science/article/pii/S0166223605001293. doi:https://doi.org/10.1016/j.tins.2005.05.004.

Gurney K, Prescott TJ, Redgrave P. A computational model of action selection in the basal ganglia. i. a new functional anatomy. Biological cybernetics

2001;84(6):401–10.

Gurney KN, Humphries MD, Redgrave P. A New Framework for Cortico-Striatal Plasticity: Behavioural Theory Meets In Vitro Data at the Reinforcement-Action Interface. PLoS Biology 2015;13(1):e1002034. doi:10.1371/journal.pbio.1002034.

Hikosaka O, Takikawa Y, Kawagoe R. Role of the basal ganglia in the control of purposive saccadic eye movements. Physiological reviews 2000;80(3):953–78.

Hong S, Hikosaka O. Dopamine-mediated learning and switching in cortico-striatal circuit explain behavioral changes in reinforcement learning. Frontiers in behavioral neuroscience 2011;5:15.

Huang C, Resnik A, Celikel T, Englitz B. Adaptive spike threshold enables robust and temporally precise neuronal encoding. PLoS computational biology 2016;12(6):e1004984.

Humphries MD, Lepora N, Wood R, Gurney K. Capturing dopaminergic modulation and bimodal membrane behaviour of striatal medium spiny neurons in accurate, reduced models. Frontiers in computational neuroscience 2009;3:26.

Humphries MD, Stewart RD, Gurney KN. A physiologically plausible model of action selection and oscillatory activity in the basal ganglia. Journal of Neuroscience 2006;26(50):12921–42.

Hunnicutt BJ, Jongbloets BC, Birdsong WT, Gertz KJ, Zhong H, Mao T. A comprehensive excitatory input map of the striatum reveals novel functional organization. eLife 2016;5:e19103. doi:10.7554/eLife.19103.

Izhikevich EM. Solving the distal reward problem through linkage of STDP and dopamine signaling. Cerebral Cortex 2007;17(10):2443–52. doi:10.1093/cercor/bhl152.

Legenstein R, Pecevski D, Maass W. A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback. PLOS Computational Biology 2008;4:1–27. URL: https://doi.org/10.1371/journal.pcbi.1000180. doi:10.1371/journal.pcbi.1000180.

Levy W, Steward O. Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. Neuroscience 1983;8(4):791–7.

Lindahl M, Kotaleski JH. Untangling basal ganglia network dynamics and function: role of dopamine depletion and inhibition investigated in a spiking network model. eneuro 2016;3(6).

Lo CC, Wang XJ. Cortico–basal ganglia circuit mechanism for a decision threshold in reaction time tasks. Nature neuroscience 2006a;9(7):956–63.

Lo CC, Wang XJ. Cortico–basal ganglia circuit mechanism for a decision threshold in reaction time tasks. Nature neuroscience 2006b;9(7):956–63.

Masquelier T, Hugues E, Deco G, Thorpe SJ. Oscillations, Phase-of-Firing Coding, and Spike Timing-Dependent Plasticity: An Efficient Learning Scheme. Journal of Neuroscience 2009;29(43):13484–93. doi:10.1523/JNEUROSCI.2207-09.2009.

McGill R, Tukey JW, Larsen WA. Variations of Box Plots. The American Statistician 1978;32(1):12–6. doi:10.2307/2683468.

McLelland D, Paulsen O. Neuronal oscillations and the rate-to-phase transform: mechanism, model and mutual information. The Journal of physiology 2009;587(4):769–85.

Miller BR, Walker AG, Shah AS, Barton SJ, Rebec GV. Dysregulated information processing by medium spiny neurons in striatum of freely behaving mouse models of huntington's disease. Journal of neurophysiology 2008;100(4):2205–16.

Mulder MJ. The temporal dynamics of evidence accumulation in the brain. Journal of Neuroscience 2014;34(42):13870–1. doi:10.1523/JNEUROSCI.3251-14.2014.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. science 2004;304(5669):452–4.

Parent A, Hazrati LN. Functional anatomy of the basal ganglia. ii. the place of subthalamic nucleus and external pallidium in basal ganglia circuitry. Brain research reviews 1995;20(1):128–54.

Potjans W, Morrison A, Diesmann M. A spiking neural network model of an actor-critic learning agent. Neural computation 2009;21(2):301–39.

Rajendran B, Sebastian A, Schmuker M, Srinivasa N, Eleftheriou E. Low-power neuromorphic hardware for signal processing applications: A review of architectural and system-level design approaches. IEEE Signal Processing Magazine 2019;36(6):97–110.

Ratcliff R, Frank MJ. Reinforcement-based decision making in corticostriatal circuits: mutual constraints by neurocomputational and diffusion models. Neural computation 2012;24(5):1186–229.

Rubin JE, Vich C, Clapp M, Noneman K, Verstynen T. The credit assignment problem in cortico-basal ganglia-thalamic networks: A review, a problem and a possible solution. European Journal of Neuroscience 2021;53(7):2234–53.

Schultz W. Dopamine signals for reward value and risk: basic and recent data. Behavioral and brain functions 2010;6(1):1–9.

Sen-Bhattacharya B, James S, Rhodes O, Sugiarto I, Rowley A, Stokes AB, Gurney K, Furber SB. Building a spiking neural network model of the basal ganglia on spinnaker. IEEE Transactions on Cognitive and Developmental Systems 2018;10(3):823–36.

Shen W, Flajolet M, Greengard P, Surmeier DJ. Dichotomous dopaminergic control of striatal synaptic plasticity. Science 2008;321(5890):848–51.

Stehman SV. Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment 1997;62(1):77–89. URL: https://www.sciencedirect.com/science/article/pii/S0034425797000837. doi:https://doi.org/10.1016/S0034-4257(97)00083-7.

Suryanarayana SM, Kotaleski JH, Grillner S, Gurney KN. Roles for globus pallidus externa revealed in a computational model of action selection in the basal ganglia. Neural Networks 2019;109:113–36. doi:10.1016/j.neunet.2018.10.003.

Sutton RS, Barto AG, Williams RJ. Reinforcement learning is direct adaptive optimal control. IEEE Control Systems Magazine 1992;12(2):19–22.

Taherkhani A, Belatreche A, Li Y, Cosma G, Maguire LP, McGinnity TM. A review of learning in biologically plausible spiking neural networks. Neural Networks 2020;122:253–72.

Tavanaei A, Ghodrati M, Kheradpisheh SR, Masquelier T, Maida A. Deep learning in spiking neural networks. Neural Networks 2019;111:47–63.

Vasilaki E, Frémaux N, Urbanczik R, Senn W, Gerstner W. Spike-based reinforcement learning in continuous state and action space: when policy gradient methods fail. PLoS computational biology 2009;5(12):e1000586.

Yagishita S, Hayashi-Takagi A, Ellis-Davies GC, Urakubo H, Ishii S, Kasai H. A critical time window for dopamine actions on the structural plasticity of dendritic spines. Science 2014;345(6204):1616–20.

5. Supplementary materials

5.1. Supplementary methods

5.1.1. Single-striatal-neuron model and experiments

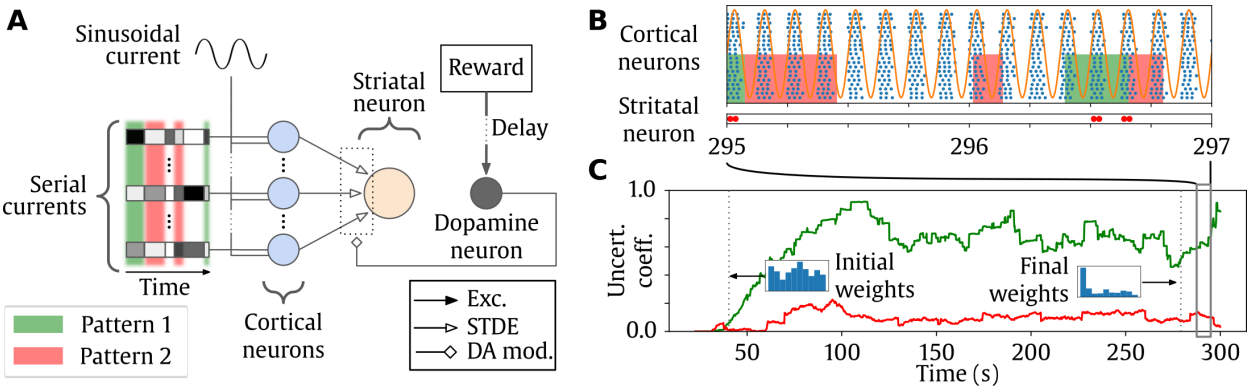

Figure 8: Pattern detection experiments with reinforcement learning and a single striatal neuron. **A.** Single-striatal-neuron model setting, with serial and oscillatory input currents feeding to a cortical layer. In this simulation, two different input patterns are used and colored in green and red. The cortical layer feeds the striatal neuron with plastic synapses with STDE, where learning occurs. A reward or punishment signal is delivered to a global dopaminergic neuron that modulates the plastic synapses. **B.** Raster plot of the cortical neurons (blue dots), with input patterns containing only half of the cortical neurons, oscillatory driving current (solid red line), and the striatal neuron (bottom, red dots). **C.** Evolution of the striatal neuron's response to each input pattern through time measured using a uncertainty coefficient (see details in the methods section). Insets show the distribution of synaptic weights at the beginning and the end of the learning procedure.

In order to assess the learning capabilities of the proposed model, we define two types of experiments: *pattern detection* and *action-selection*. The latter one is already explained in the main text. During pattern detection experiments (Supplementary Fig. 8B), we train a simple model to detect one specific pattern within a noisy input stream. Two (the so-called selected and non-selected) non-temporally-overlapping repeating patterns are presented 20% of the time each (40% in total). We test the STDE learning rule in a RL setting, where a reward (excitation to the dopaminergic neuron) is given if the striatal neuron spikes sometime after the selected pattern is presented. Otherwise, if the striatal neuron fires in response to the non-selected pattern, punishment (inhibition to the dopaminergic neuron) is given to the striatal neuron. Finally, as a stress test, we added a policy swapping procedure for switching the rewarded pattern every 200 seconds (Supplementary Fig. 9). This way, we can test how robust is our combination of synaptic and homeostatic rules during learning.

For this first set of experiments, we used a model with only one striatal neuron that learns to solve a simple RL task. This model allows the validation of the proposed learning mechanisms. It uses the input protocol explained in the oscillatory drive section 2.1.2. A dopaminergic signal modulates the synapses that connect from the cortical neurons to the striatal neuron (Supplementary Fig. 8A), implementing the STDE learning rule as well as the homeostatic mechanisms. Rewards (punishments) delivered by the environment alter the dopaminergic modulatory signal by exciting (inhibiting) the dopaminergic neuron every time the striatal neuron spikes when the input pattern is correct (incorrect). The environment delivers rewards and punishments with some delay (fixed to 300 ms by default). If the striatal neuron does not fire, the environment delivers no reward nor punishment to the DA neuron.

5.1.2. Mutual information

In order to measure how good the detection is in the pattern detection experiments, we calculated the mutual information (MI) between the presentation of each input pattern and the striatal neuron activity, as previously done in Garrido et al. (2016). We consider that the striatal neuron responded to the pattern if it fires at least once during the stimulus presentation, lasting from 100 to 500 ms following a uniformly distributed random distribution. For each stimulus used in the pattern detection experiments, we consider the possible states S of the pattern (present or absent) and the possible response R of the striatal (neuron fired or not). The MI is then defined in equation (6).

$$MI = H(S) + H(R) - H(S, R) \tag{6}$$

where $H(S)$ is the entropy of the stimuli patterns, $H(R)$ is the entropy of the responses, and $H(S, R)$ the joint entropy of the stimuli patterns and the responses. These values are defined as in Garrido et al. (2016). The upper bound of the MI for a perfect detector would be $MI_{max} = H(S)$, so we can obtain a normalized measurement of performance called uncertainty coefficient (UC) defined in equation (7). The UC is calculated independently for both the rewarded and the non-rewarded patterns during pattern detection experiments.

$$UC = \frac{MI}{MI_{max}} = \frac{H(S) + H(R) - H(S, R)}{H(S)} \tag{7}$$

5.1.3. Parameters used

Parameter	Cortical	Striatal	Action	Dopaminergic
e_{exc} (mV)	0.0	0.0	0.0	0.0
e_{inh} (mV)	−85.0	−85.0	−85.0	−85.0
τ_{AMPA} (ms)	5.0	5.0	5.0	5.0
τ_{GABA} (ms)	10.0	30.0	60.0	10.0
τ_{ref} (ms)	1.0	15.0	15.0	1.0
C_m (pF)	250.0	50.0	100.0	250.0
g_{leak} (nS)	25.0	10.0	25.0	25.0
V_{thr} (mV)	−40.0	−50.0	−40.0	−65.0
e_{leak} (mV)	−65.0	−65.0	−65.0	−40.0

Table 1: Neuron parameters used in the model.

MSN D1		MSN D2	
Parameter	Value	Parameter	Value
k^-_{lo}	0.0	k^-_{lo}	−1.0
k^+_{lo}	−1.0	k^+_{lo}	1.0
k^-_{hi}	−1.0	k^-_{hi}	0.0
k^+_{hi}	1.0	k^+_{hi}	−1.0

Table 2: STDE parameters used in the model.

5.2. Supplementary results

5.2.1. Single-striatal-neuron experiments

In a previous article by Masquelier et al. (2009), an oscillatory driving signal greatly facilitates the recognition of complex patterns over noise with STDP-like rules. We have extended this learning rule to account for a rewarding signal in a RL paradigm. During a whole learning task (lasting 200 seconds), two different repeating and non-overlapping input patterns are presented. Only one of them produces a rewarding signal if, and only if, the striatal neuron fires simultaneously to the pattern presentation, providing reward modulation to the learning rule. Using this RL framework, the striatal neuron becomes selective to the presentation of the rewarded pattern only (Supplementary Fig. 8B). It usually takes less than 100 seconds of simulated time to consistently generate spikes with the presentation of the rewarded pattern (Fig. 8B). The detection capabilities of this network are also evidenced by the evolution of the uncertainty coefficient (green line in Supplementary Fig. 8C), which remains stable between 0.6 and 0.8 after 80 seconds of discontinuous pattern presentation (Supplementary Fig. 8C), while the punished pattern receives no considerable response (red line in Supplementary Fig. 8C). It can also be observed how the initial uniform weight distribution (insets in Supplementary Fig. 8C) turns into a binomial distribution with a small number of synapses with near-maximum weights and most of the synapses near the minimum weight.

Once demonstrated the effectiveness of the STDE learning rule, we aim to assess if it allows detection of rewarded patterns with policy swapping (i.e., the pattern that offers rewarding signals is swapped every 400 seconds of simulation). Every time that the rewarding policy swaps, the neuron temporarily reduces its average firing rate (cyan line in Supplementary Fig. 9Di), and consequently, the adaptive firing threshold approaches the resting potential (pink line in Supplementary Fig. 9Di). Once the threshold is low enough, the neuron starts learning the new rewarded pattern, increasing the activity of the dopaminergic neuron as a consequence (Supplementary Fig. 9Ei). This is an important feature because neurons can recover from silent states caused by sudden changes in the reward policy.

Inspired by the different types of neurons existing in the striatum, we adapted the synaptic model parameters to reproduce the differential operation of the learning rule for the MSN D1 and the MSN D2 neurons (MSN D1 and D2 parameters for STDE in Supplementary Table 2). Thus, we adjusted different kernel shapes for low and high DA (Supplementary Fig. 9Bi and Bii, left and right, respectively). According to our simulations, the neuron equipped with a D1 kernel learns to detect only the rewarded pattern (Supplementary Fig. 9Ci). In contrast, the striatal neuron equipped with MSN D2 kernel parameters (a reversed version of MSN D1) learns to detect the non-rewarded pattern (Supplementary Figs. 9Cii, 9Dii and 9Eii). These results point out that, in a network of MSNs with D1 and D2 subpopulations, the D1 subpopulation learns to respond to rewarded patterns while the D2 neurons learn to fire in response to the punished (or non-rewarded) patterns. In this way, the output layer makes simple decisions by just weighting the activity of these subpopulations.

5.2.2. Homeostatic mechanisms: non-Hebbian strengthening and adaptive threshold

Aiming to check the influence of the homeostatic mechanisms, we have replicated the same policy-swapping learning framework with a more complex task (five different input patterns) and different configurations of the homeostatic rules. In the absence of non-Hebbian strengthening, successful learning requires fine-tuning of the learning rule parameters and maximum weight for

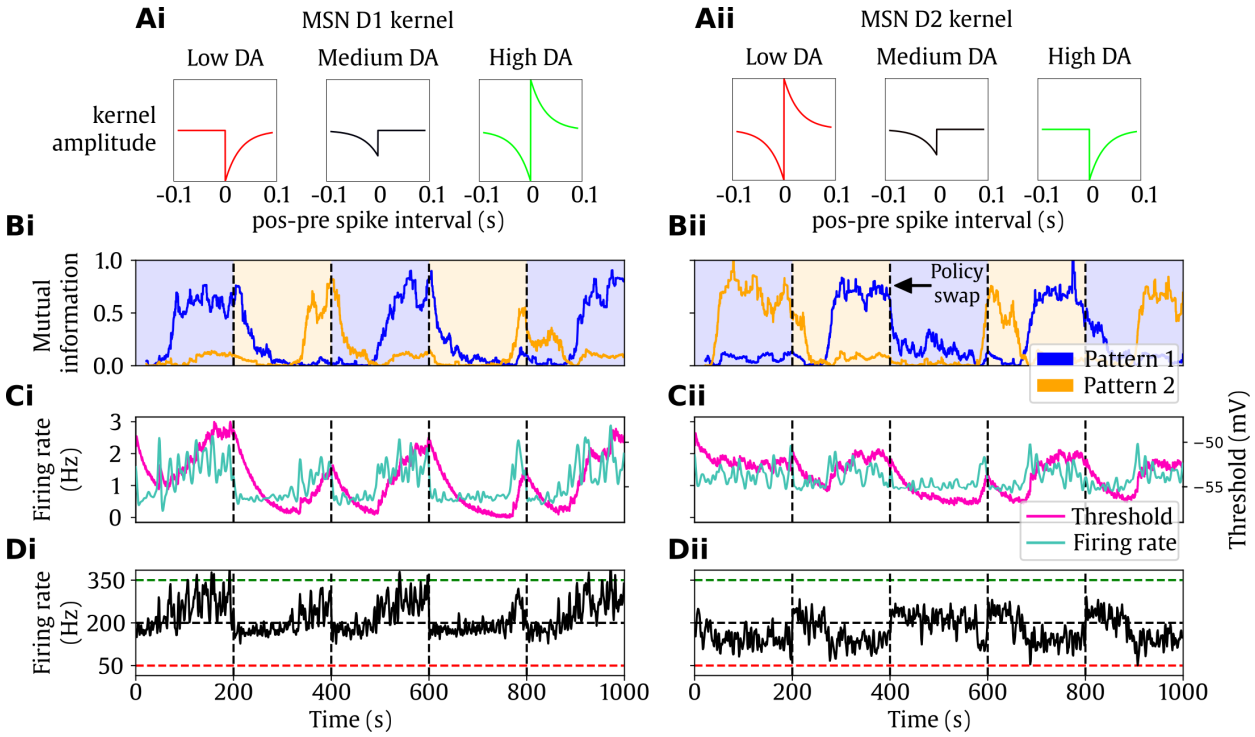

Figure 9: Pattern detection experiments with two different STDE sets of parameters. **Xi** column shows the training resulting from using a learning kernel adapted to learn rewarded patterns, as used in MSN D1 synapses. **Xii** column shows the same training results for the MSN D2 kernel used. Note that this kernel is learning the opposite (punished) pattern. **A** row shows the kernel functions used with different levels of DA. **B** row shows the response of the striatal neuron. The background color indicates which pattern is being rewarded at that specific time frame, and the vertical dotted lines indicate when the rewarding policy swaps. **C** row shows the evolution of the adaptive threshold and the firing rate of the striatal neuron. **D** row shows the firing rate of the dopaminergic neuron, which represents the amount of reward obtained by the striatal neuron through the task. The horizontal green, black and red dotted lines indicate the maximum, baseline, and minimum dopaminergic activity.

each simulation seed (data not shown). Thus, we barely managed to find a set of parameters suitable for multiple seeds without this homeostatic mechanism. For this reason, in all the simulations shown in this article we employ the non-Hebbian strengthening mechanism.

On the other hand, the adaptive threshold is not strictly necessary for successful learning. However, the learning performance (in terms of UC) with adaptive threshold increases faster and more reliably than without adaptive threshold (Supplementary Fig. 10). It is important to highlight that lack of homeostatic mechanisms often resulted in the more frequent inability of detecting cortical patterns, as demonstrated by lower MI values for the 25-percentile of the simulations (lower boundary of blue shadow in Supplementary Fig. 10, right). In the absence of these mechanisms, the striatal neuron activity extinguishes when the reinforcement policy swaps and, in many cases, remain silent for the rest of the simulation. We tested different learning rates and time constants of dopamine and, in every case, learning was faster with adaptive threshold, as shown in Supplementary Fig. 14. Thus, these homeostatic rules provide the STDE rule with the ability to re-learn different patterns reliably. Moreover, using both of these mechanisms also makes learning robust within a broader parameter space and makes it unnecessary to fine-tune the parameters for each experiment. Although only one of these homeostatic mechanisms would be enough to avoid silent

neurons, we saw in our tests that the system recovered faster and more reliably by using both.

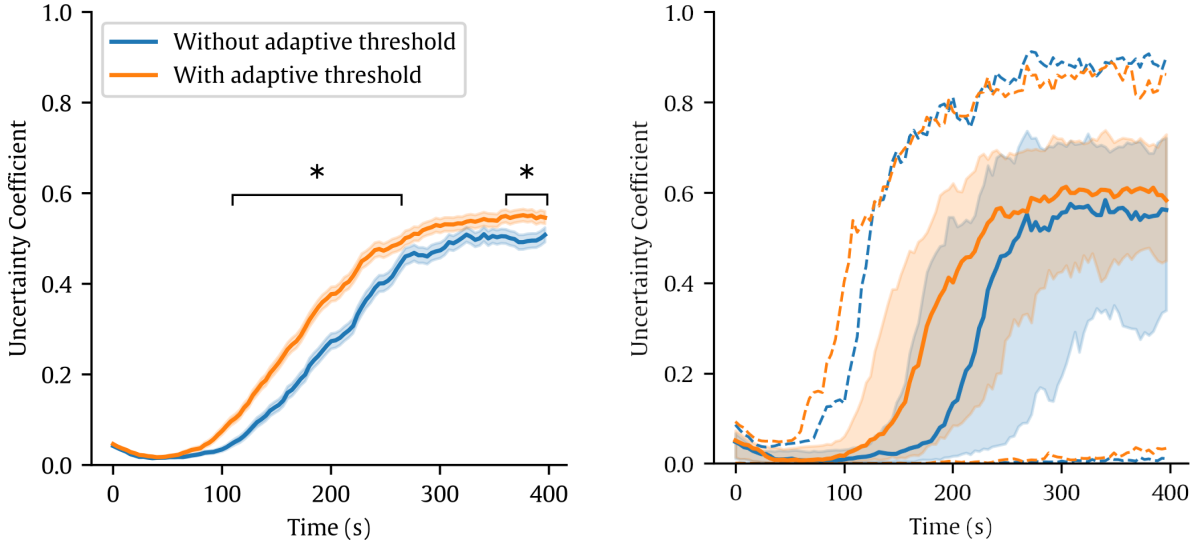

Figure 10: Effect of the adaptive threshold in the learning performance of the single-striatal-neuron model. In this experiment we used a more complex version of the policy-swap task with 5 (one rewarded, the rest punished) different patterns instead of 2. The left curves and the filling represent the evolution of the mean uncertainty coefficient and standard error during a repeated 400-s learning protocol (n=300). The asterisks marked intervals indicate where the means are statistically different with 95% confidence level. The right plot shows the percentiles 5-25-50-75-95, with dashed lines (5 and 95 percentiles), fillings (25 and 75 percentiles) and solid lines (50 percentiles).

5.2.3. Effect of reward delay and input pattern

We wondered how the delay between the action decision (in response to cortical stimulus) and the rewarding signal affects the learning capabilities of our system. In order to evaluate the impact of this parameter, we carried out network simulations with different reward delays (we did not have to adjust any other parameter due to the robustness of the model). We found the best performance when the rewarding signal was provided 300 ms after the sensorial presentation (blue line in Supplementary Fig. 11). Longer or shorter delays resulted in decaying learning accuracy. This result is similar to what can be found in biology ((Yagishita et al., 2014)).

Since our implementation of the DA-modulated learning rule is based on eligibility traces, we wondered if this optimal delay was somehow related to the duration of the stimulation patterns. Then, we evaluated the reward delay effect on learning when sensorial patterns were longer (300-700 ms and 500-900 ms) than in the control case (100-500 ms). However, our simulations show similar learning accuracy with longer cortical patterns (orange and green lines in Supplementary Fig. 11) as in control conditions (blue line in Supplementary Fig. 11). So that it seems unlikely that the pattern generation algorithm influenced the preferred delay.

Finally, we also studied how the frequency of pattern presentation influences the accuracy achieved at the end of the simulation. We compared the results obtained presenting the patterns 80 percent of the time (as in the rest of the experiments made) with the results obtained by presenting the patterns 40 percent. In order to compensate for the lower exposure of the striatal neurons to input patterns (since in the latter, the network will only see the patterns half the time), we simulated twice as long (up to 1000 seconds). According to our simulations, the proposed network similarly managed to successfully associate cortical inputs to associated actions independently

of how often the patterns are presented, as long as it experiences enough trials (Supplementary Fig. 12).

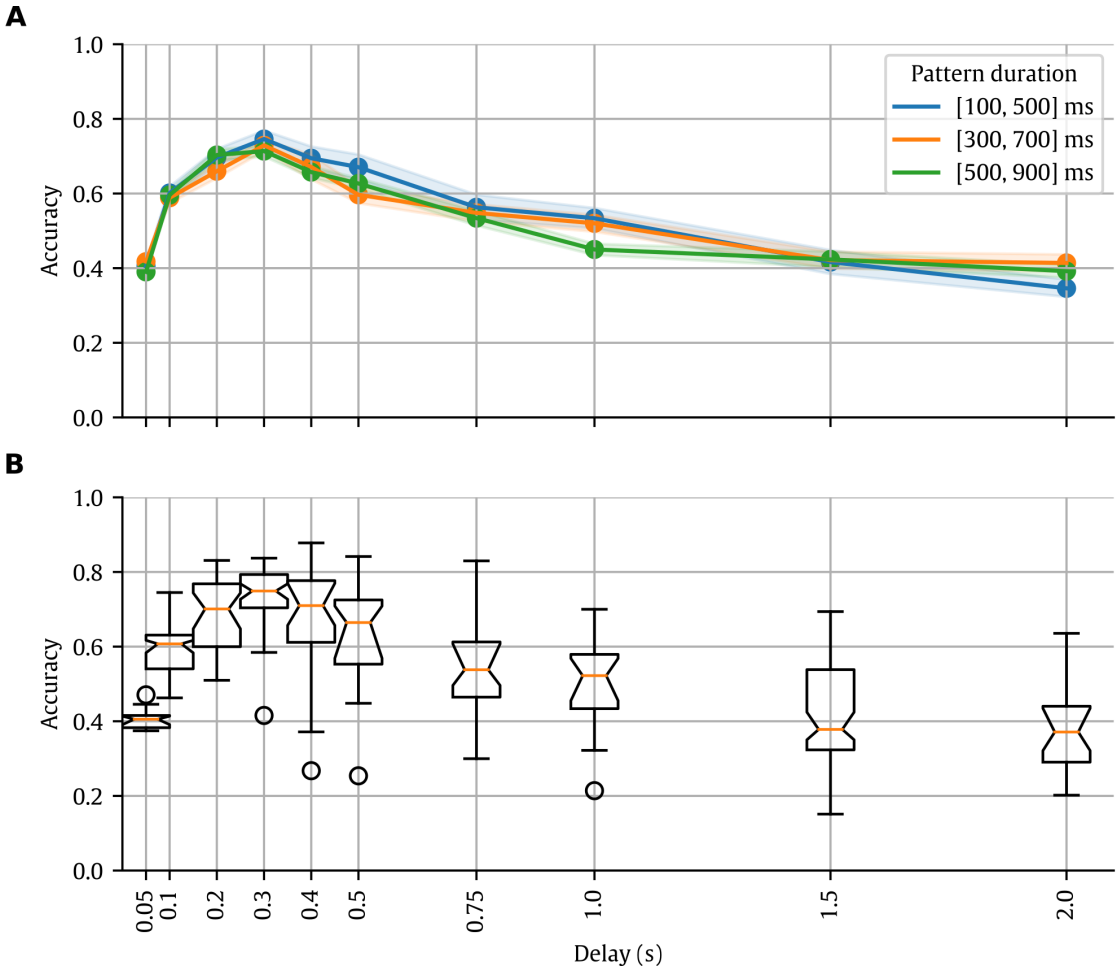

Figure 11: Effect of the delay of the rewarding feedback in the learning accuracy. **A**. Simulations with different pattern lengths: within 100 to 500 ms (blue), within 300 to 700 ms (in orange), and within 500 to 900 ms (in green). Every point represents the mean accuracy level obtained in the last 100 seconds of simulation with different delay values, and the shaded area shows the standard error of the mean (n=10). **B**. Notched box plot of all the values. Notice that "if the notches about two median do not overlap, the medians are, roughly, significantly different at about a 95% confidence level" (see McGill et al. (1978) for details).

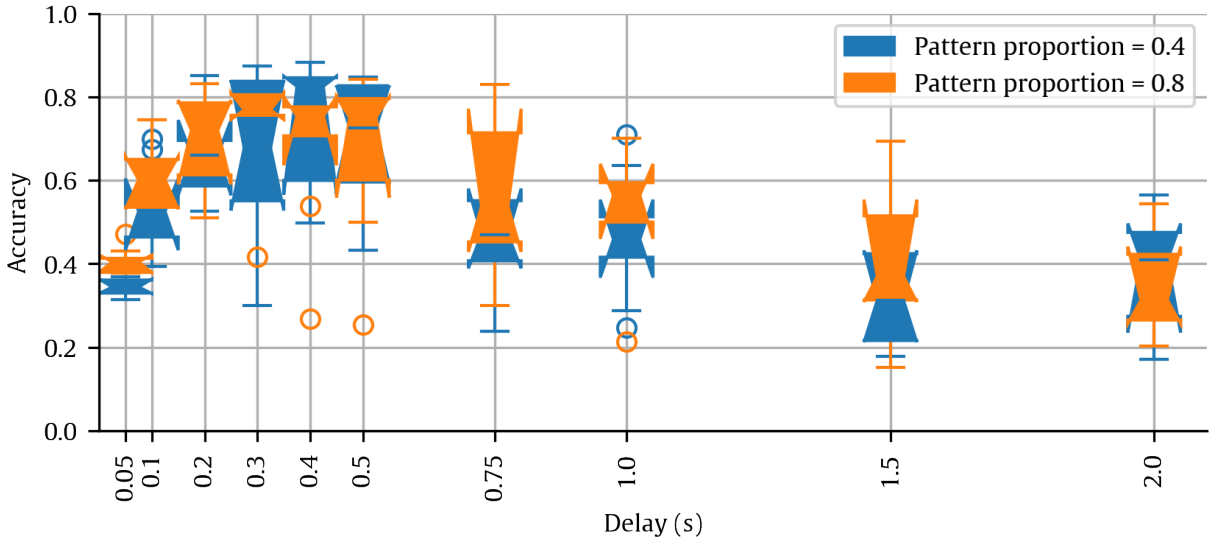

Figure 12: Effect of the delay of the reward in the learning performance with input pattern proportion of 0.8 (in blue), and with input pattern proportion of 0.4 (in orange). Every notched box (McGill et al., 1978) represents the median (n=10) performance level obtained in the last 100 seconds of simulation for different delay values.

5.2.4. Effect of lateral inhibition in harder experimental setting

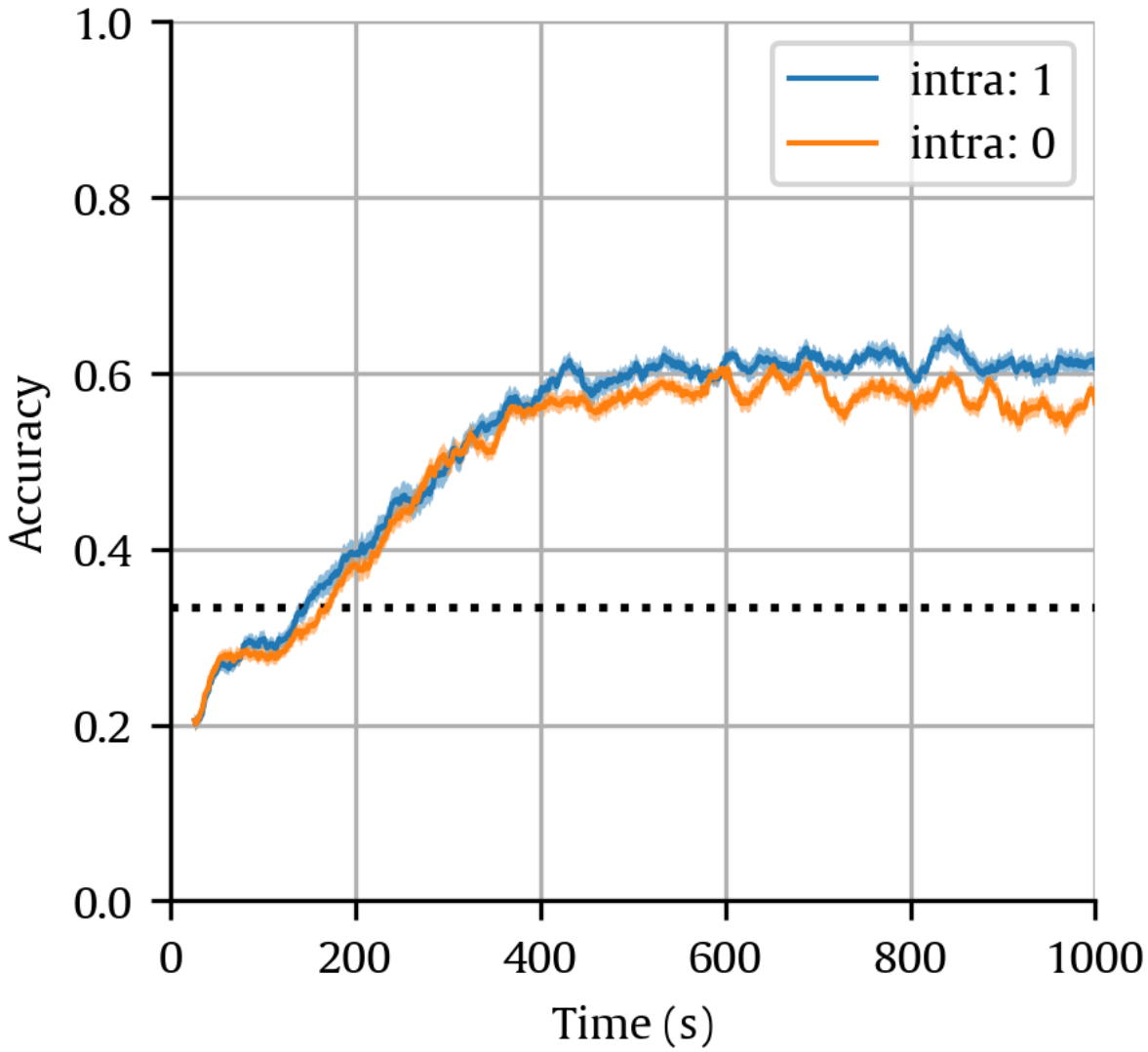

Figure 13: Same as in Fig. 6, but simulated for 1000 seconds.

5.2.5. Effect of DA time constant, learning rate and adaptive threshold

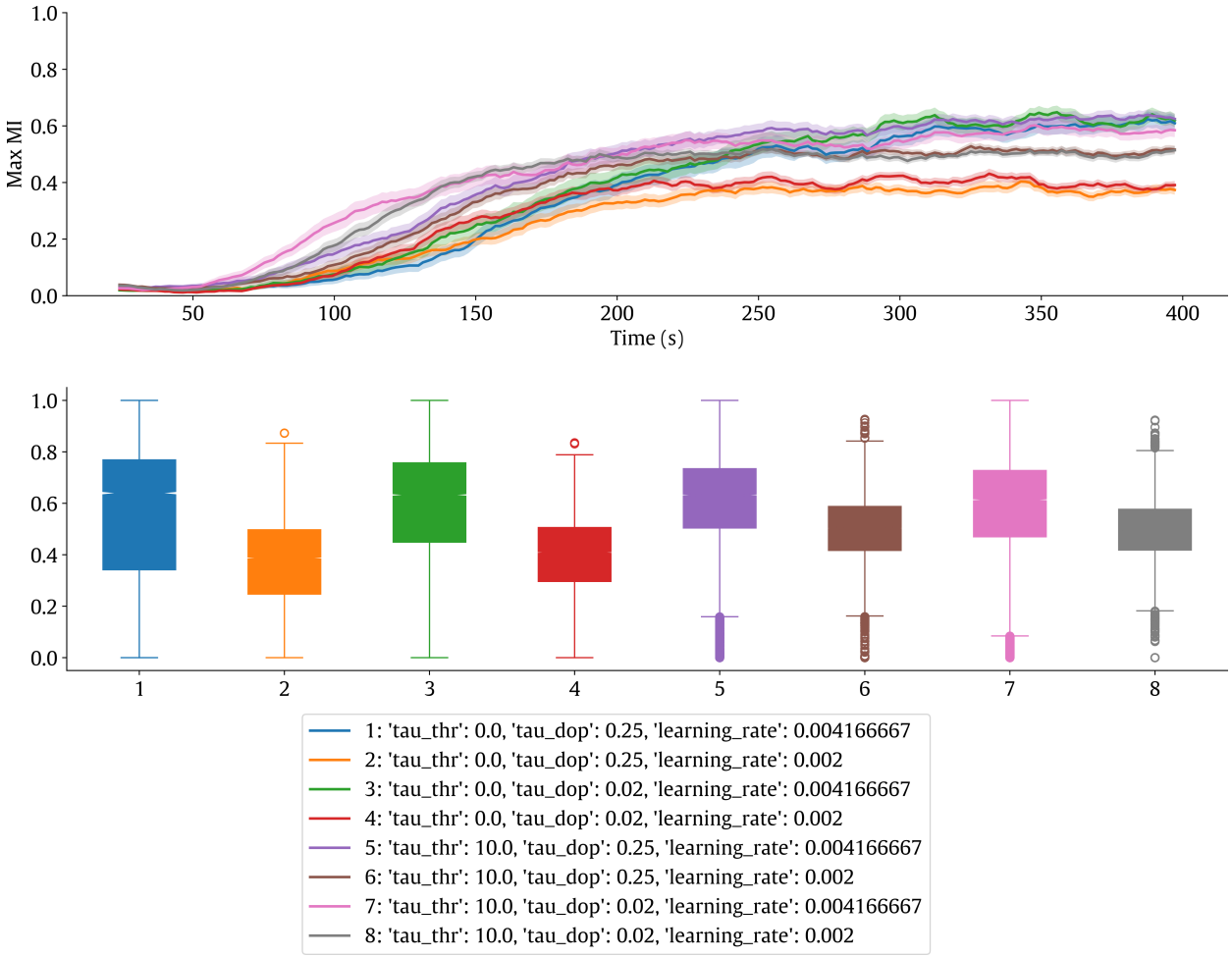

Figure 14: Learning performance for different values of DA time constant, learning rate and adaptive threshold. At the top, mean and standard error are shown for each condition. At the bottom, boxplots of the last 200 seconds of simulation (n=80).

5.2.6. *Lateral connectivity patterns effect*

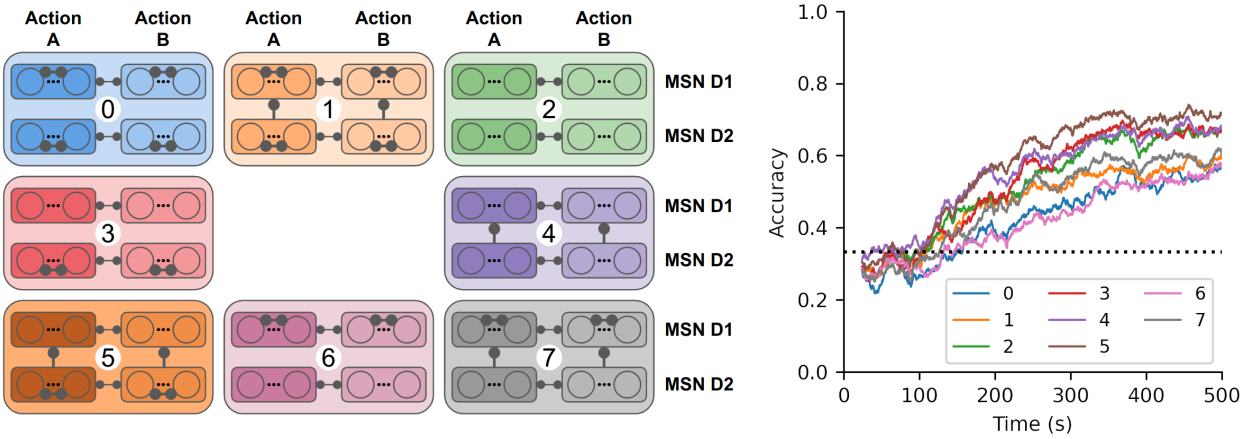

Figure 15: Learning performance for different connectivity patterns of lateral inhibition. Left: Connectivity topologies tested in these experiments. Note that all these tests assume inter-channel inhibition, as they clearly outperformed other models. Right: evolution of the learning accuracy during 500s of simulation with the medium-complexity task. Every line is marked with the same color of the topology under test. Each line represents the average value with $n = 10$ seeds