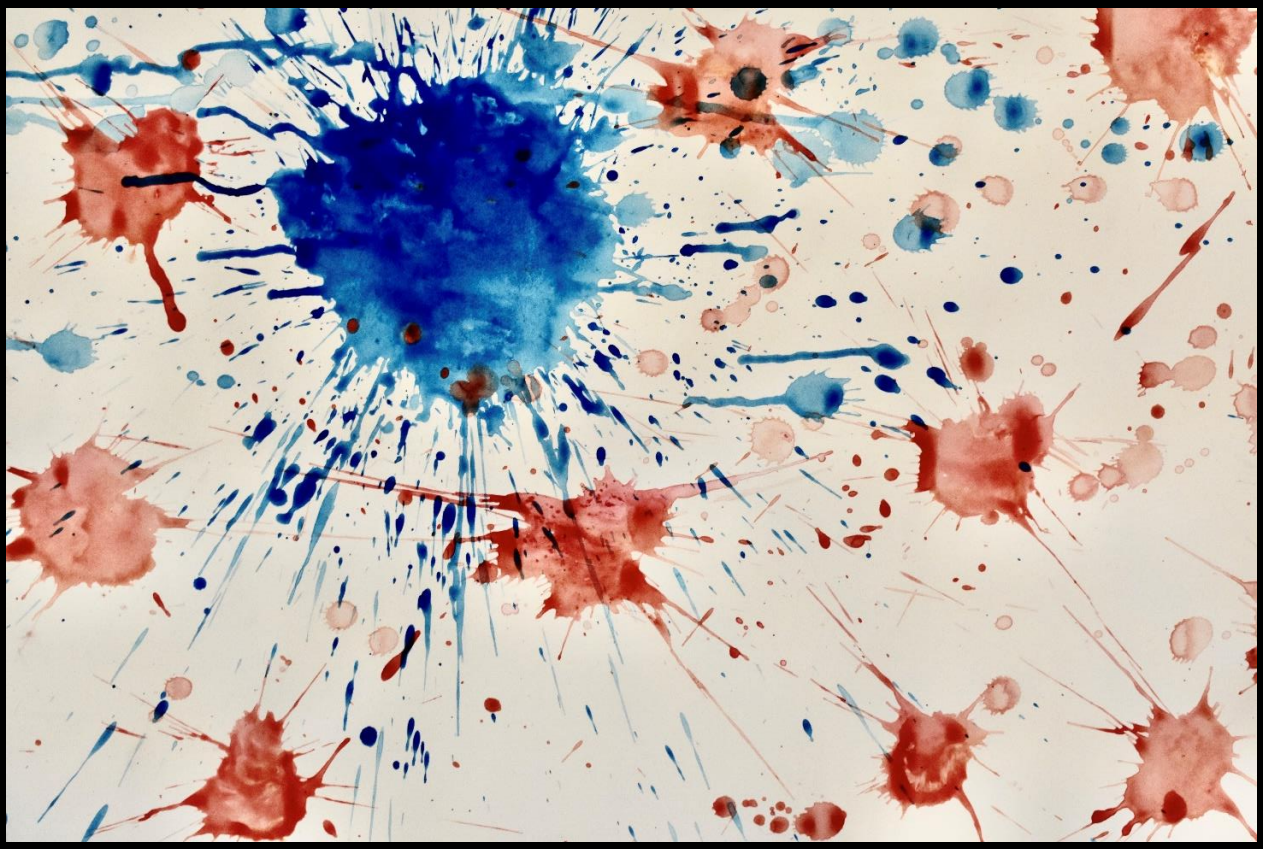
Striatal circuits supporting action production and suppression

Bruno Filipe Pereira da Cruz



Dissertation presented to obtain the
Ph.D degree in Neuroscience

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,
January, 2022

UNIVERSIDADE NOVA DE LISBOA

Striatal circuits supporting action production and suppression

Bruno Filipe Pereira da Cruz

Dissertation presented to obtain the
Ph.D degree in Neuroscience

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Research work coordinated by:

Champalimaud Foundation

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Oeiras, January, 2022

ITQB UNIVERSIDADE NOVA DE LISBOA

Striatal circuits supporting action production and suppression

Bruno Filipe Pereira da Cruz

A Dissertation
Presented to the Faculty
of Universidade Nova de Lisboa
in Candidacy for the Degree
of Doctor of Philosophy

Supervisor: Joseph J. Paton

January 2022

Cover Image
"Direct and Indirect pathway medium spiny neurons"
Painted and photographed by Benjamin Zarov

You have brains in your head.
You have feet in your shoes.
You can steer yourself
in any direction you choose.
You're on your own, and you know what you know.
And you are the guy who'll decide where to go.

You'll look up and down the streets. Look 'em over with care.
About some you will say, "I don't choose to go there."
With your head full of brains and your shoes full of feet,
you're too smart to go down any not-so-good street.

Dr. Seuss, Oh, The Places You'll Go! (1990)

Para a minha 'Bó Isabel,

Acknowledgements

As I am sitting on my chair, writing this section, I look back to when I started this journey and I can barely recognize the person I was then. A kaleidoscope of emotions and unforgettable experiences pushed me to grow scientifically but also as a human being. Such growth was not least thanks to the people I met, learned from, befriended, and loved along the way. I don't tell them often enough how profoundly grateful I am to have them in my life, and how it is in large part thanks to them that I get to write these words today.

First, I would like to thank Joe. Thanks for the guidance, for putting up with my chronic pessimism, for sharing the ups and downs of this journey, and for the support and advice throughout the years. Whatever comes next, I hope to always be able to count on you as a mentor and a friend. Thank you!

I would also like to acknowledge the members of the Paton lab (former and current). I thank them for creating a space where I was happy to come back to, day after day. A special thanks to Tiago, Filipe, Teresa, Margarida S., and Ben for feedback on different versions of this dissertation.

My INDP classmates. I thank them, specially Inês, Tati, and Paco for sharing so much of this adventure with me.

The members of my thesis committee, Megan and Leopoldo. Thank you for keeping me on track and for the help bringing this doctorate to fruition. Moreover, I must also explicitly thank them for the opportunity to TA in their respective courses. Those were some of the best weeks of my Ph.D. journey.

My gratitude also goes to the guys at the Hardware platform, Paulo, Filipe, and Dário. They "adopted" and offered me a second lab. I thank them for the technical help, discussions, lunches, all the fun projects, for entertaining my "ideas", and for their friendship. Finally, I must single out Artur. Not only is he one of the most talented and hardworking people I had the pleasure to meet and work with during my time at CCU, but he is also one of the most generous.

As a community in CCU, what we have is something very unique, special, and, I am afraid, not easily reproducible elsewhere. Many people in the institute influenced this work, directly or indirectly. They offered technical help, endless hours of discussions and made my experience in Lisbon all the more enjoyable. My gratitude goes to Ben (thanks for the cover too!), Teresa, Cindy, Joaquim, Jovin, Madalena, Dana, Andreas, Luís, Andreia, Ana, Tomás, Vicky, Pietro, and many others along the way that certainly deserve to be on this list.

Pedro and Joana, thank you for keeping me sane, especially in this final year.

Thanks to Filipe for always being there for a quick discussion and for being someone that I look up to, who embodies scientific curiosity and creativity.

A special thanks to Cristina. She was one of the smartest and kindest people I have ever come across. She challenged my worldview in many aspects, and I am so lucky to have had the opportunity to meet her.

Hugo, Sam, Vivek and Tiago. In their unique ways, they embody the best qualities I learned to value in friends. Thank you for being there for the fun, the trips, but also when I needed to talk about the frustrations of life and the lab, for providing advice, and for making life so much better.

Margarida Sousa and Margarida Pexirra, two of the brightest and kindest friends I met during my time in Lisbon. Thank you, M.S. for entertaining my ideas, for the help in the lab, and for being a good listener. Thank you M.P. for the occasional "bullying" when I needed it, and for teaching me to not take myself too seriously.

I thank Antonia for the conversations about everything and nothing, especially those over tea or ice cream!

To Cat, I am grateful for the time, the companionship, for presenting me with a different way to live life, and for showing me what I often failed to see on my own.

A special thanks to Bass. He was one of the first friends I made in CCU and has remained one of the best. I thank him for the technical help, discussions, and for the contagious dedication and energy he shared with me in my early years in Lisbon. More importantly, I am grateful for his ability to always challenge me to see things from a different angle, for being a good listener, but also, at times, for saying what needed to be said. Above all, thanks for being (,) Bass!

Sofia, my "big sister". Thank you for taking me under your wing when I joined the lab, for being there for the science, the parties, the silliness, the seriousness, the smiles, and the tears. I am grateful for our friendship.

Finalmente, para a minha família, especialmente os meus pais: Lina e Armando. Obrigado por estarem sempre presentes e pelo amor incondicional. Por serem um porto de abrigo, de repouso, de concentração, onde, por alguns dias, os problemas deixam de existir e consigo novamente ouvir os meus pensamentos. Finalmente, à minha avó, Isabel, por ser uma fonte de carinho e amor infinito. Um exemplo de bondade e caráter. A ela, dedico esta tese.

Título

Circuitos do corpo estriado suportam produção e supressão de acção

Resumo

O comportamento animal requer a combinação de processos de supressão e produção de ações. Um predador deve suprimir o ataque à sua presa até que esta esteja ao seu alcance, assim como nós, humanos, temos que, por vezes, não ceder à tentação do momento a fim de alcançar maiores ganhos no futuro.

Perturbações do equilíbrio destes dois processos são prevalentes em estados patológicos dos gânglios da base (BG), tal como: perturbação de hiperatividade/défice de atenção (PHDA), doença de Parkinson e doença de Huntington.

Os dois principais circuitos dos BG, a via direta e indireta, são classicamente associados à estimulação e supressão de ações, respectivamente. Todavia, estudos mais recentes reportam que, durante o movimento, populações de neurónios principais do corpo estriado, pertencentes à via direta (dMSN) ou indireta (iMSN), exibem co-activação. Esta observação levou muitos a questionar o modelo clássico de oposição funcional destes dois circuitos.

Nesta tese é mostrado que, não obstante da observação de co-activação alinhada a movimento, durante um período dominado por supressão de ação é possível identificar sinais de oposição funcional entre as duas vias. Primeiro, medindo a atividade das populações neuronais das duas vias, em áreas dorsolaterais do corpo estriado (DLS) de roedores, mostramos que a atividade de iMSN, durante um período de supressão de ação, é comparativamente superior à dos dMSN. Mais, os padrões de atividade, em cada um dos hemisférios monitorizados, variam em direções opostas, e nas duas vias, qualitativamente representando a necessidade de suprimir uma suposta ação contralateral a ser planeada.

Consistente com estas observações, experiências em que a actividade neuronal de neurónios iMSN em DLS foi silenciada, revelam que esta população é necessária para determinar *se*, *quando* e *quais* ações que os animais são capazes de suprimir.

Por outro lado, manipulação dos neurônios da via direta (dMSN) em DLS não produziu um efeito detectável na habilidade dos animais para suprimir acções, mas sim no vigor com que estes as produzem.

Tendo em conta estas observações, assumimos que o recrutamento da via indireta (iMSN DLS), durante a tarefa de comportamento analisada nesta tese, surgiu da necessidade de contrariar um sinal "positivo" para agir, proveniente num circuito putativamente adjacente. Inspirados por algoritmos de aprendizagem de inteligência artificial (*Reinforcement Learning*), pensados ser parcialmente implementados no circuito dos BG, construímos um modelo computacional em que múltiplos agentes paralelos são combinados para gerar decisões. Depois de treinado numa tarefa idêntica à usada em experiências animais, o modelo é capaz de reproduzir os dados de comportamento animal, assim como os resultados da monitorização das dinâmicas e manipulações neuronais. Adicionalmente, obedecendo a uma previsão do modelo, inibição optogenética da via direta em circuitos dorso-mediais dos BG (DMS) produziu efeitos consistentes com este circuito providenciar o sinal "positivo" que o DLS aprende a contrariar.

O trabalho apresentado nesta tese, em adição a reconciliar observações prévias relativas à normal função dos BG, propõe um possível modo de operação no qual múltiplos circuitos paralelos dos BG interagem a fim de gerar comportamento animal.

Abstract

Adaptive behavior involves a judicious combination of suppression and production of actions. A predator must suppress its urge to pounce until its prey is within reach, just as humans must suppress giving in to temptation to secure longer-term rewards. Imbalances between these two processes are associated with disorders of the basal ganglia (BG) such as ADHD, Parkinson's, and Huntington's diseases.

The direct and indirect pathways of the BG, were classically thought to promote and suppress action, respectively. However, observed coactivation of neurons initiating the two pathways, striatal direct (dMSNs) and indirect (iMSNs) medium spiny neurons, has led many to question that view. In this dissertation, I will show that while coactivation occurred during movement, action suppression produced systematic features of functional opponency between the pathways. Fiber photometry in the dorsolateral striatum (DLS) of mice revealed relatively higher levels of iMSN activity during a period of action suppression when compared to dMSNs. Furthermore, relative activity in the two hemispheres evolved over time and in opposite patterns across the two pathways, qualitatively following the need to suppress a planned contra-lateral action. Consistent with these observations, optogenetic inhibition of DLS iMSNs disrupted *whether*, *when* and *which* actions mice were able to suppress, whereas inhibition of DLS dMSNs had no detectable effect on action suppression or selection but was instead required for movement invigoration.

iMSN DLS engagement was reasoned to arise to counteract a drive to act, located elsewhere. By constructing a simplified multi-agent reinforcement learning model we were able to reproduce behavior, physiological and manipulation results in DLS. Additionally, inhibiting dorsomedial striatal (DMS) dMSN population produced qualitatively similar effects on choice behavior as those predicted by the model when inhibiting the sub-agent responsible for promoting action, suggesting that, in our task, associative striatal circuits are part of the circuitry responsible for carrying the drive to act.

In addition to reconciling previous observations of striatal function, our study highlights how interaction between multiple parallel processes can lead to effective behavior control by the brain.

Author Contributions

Chapters 1 and 3 were written by Bruno Cruz. Experiments in chapter 2 were designed by Bruno Cruz and Joseph Paton. Experimental data were acquired by Bruno Cruz. Ben Zarov and Daniela Domingues assisted with the training of some of the animals included in the dataset. Data analysis was carried out by Bruno Cruz with assistance and supervision from Joseph Paton. The computational reinforcement learning model presented in chapter 2 was conceived by Gonçalo Guiomar, Bruno Cruz and Joseph Paton. Gonçalo Guiomar implemented the model and performed simulations. Chapter 2 is adapted from a pre-print manuscript (Cruz et al., 2020) written by Bruno Cruz and Joseph Paton.

Financial Support

This work was funded by Fundação para a Ciência e a Tecnologia (FCT) and Fundação Champalimaud. Bruno Cruz is a student of the International Neuroscience Doctoral Program (INDP) 2014 and was supported by the Ph.D. fellowship PD/BD/105945/2014 from FCT.

Overview

This thesis is composed of three main chapters. In the first chapter, I present a general introduction where a review of the literature relevant for Chapters 2 and 3 is made. In this introduction, I start by framing the general problem of *action selection* that biological agents are faced with and how *action suppression* plays a pivotal role in its solution. The remaining of this first chapter is dedicated to reviewing how a set of brain areas, so-called basal ganglia, function to support the process of action production and suppression. I will start with a short account of the basic functional anatomy of the circuit, its general organizing principles, and relevant models of its function.

In Chapter 2 I present the main findings of my doctoral work. In it, I will describe how we combined a novel behavioral paradigm, designed to study action production and suppression in mice, with cell-type-specific activity recordings and manipulations to probe the functional role of the direct and indirect pathways of the basal ganglia.

The monograph will conclude with a third final chapter where I discuss the general implications of our findings to the field, along with remaining open questions.

Contents

Chapter 1

General introduction

1.1 Act First

Organisms evolved to interact with the world. While senses of seeing, hearing, smelling, etc., are certainly useful in that they endow a system with the ability to *feel* its surroundings, it presents little to no ethological advantage if the same biological system is not afforded the ability to act on the environment. Moreover, *acting* alone is not enough, and along with it organisms likely evolved the ability to act in ways that are, ultimately, beneficial for them. This is an ubiquitous truth, from simple unicellular organisms that use a molecular gradient to orient in the hopes of finding a better source of energy substrate, to Human beings writing a Ph.D. thesis attempting to graduate.

Some interactions, like absentmindedly withdrawing our hand from a scorching flame, are reflexive and require little or no thought. Others, often call for the representation of an explicit goal, deliberation among several possible options, or might require learning and modification throughout the animal's life. Such volitional, self-generated behaviors, must leverage control algorithms that are capable of successful and efficiently regulating animal behavior contingent on its internal state (*e.g.,* hunger, thirst), sensory input relaying contextual information from the environment (*e.g.,* existence of a predator), past experience and available action repertoire.

We often take for granted such ability to will and execute volitional behavior control by regulating a presumed action selection (AS) process, yet several central nervous system pathologies are characterized by a deep and debilitating inability to do so. Furthermore, pathological behavior observed in many of these conditions appears to stem from the inability to successfully suppress unsuitable actions, rather than to produce the correct one (*e.g.,* addiction).

In this dissertation, I will argue that a set of vertebrate structures, the basal ganglia (BG), support the selection of appropriate actions, depending on expected outcomes, by simultaneously regulating the processes of action production and suppression.

1.1.1 The action selection problem

Adaptive autonomous agents, be them biological or artificial, inhabit environments in which they must satisfy a set of internal drives, goals and motivations. In order to achieve these goals, most agents possess the ability to collect information from the world, through sensors, and in turn, to interact with it by performing *actions* through available effectors (*e.g.,* arms, legs, wheels, *etc.*). In the course of behavior, one is able to intuitively identify clear inflection points that correspond to transitions from one behavior to another, wherein animals must decide *"what to do next"*. Picture a gazelle grazing the savanna. The same visual scene input, collected through its sensory organ, might carry simultaneous information on a new source of plentiful food as well as a potentially camouflaged predator. What should our agent do? On one hand, immediate survival should take precedent and the gazelle should run away from the hypothetical predator. On the other hand, if the gazelle has not eaten in days, a new patch of food becomes quite tempting, perhaps taking precedent. On top of that, the predator might be hiding, making its detection difficult. The gazelle might then decide to risk it and approach the area. I believe this scenario exemplifies rather well the ethologically relevant decision making demands biological agents are faced with. Generally speaking, given the multitude of interval drives and time-varying goals animals are inherently bound to, the space of potential actions to be performed in order to satisfy these drives is, at any given time, large.

In theory, multiple drive→action channels could be pursued in parallel. However, agents, especially biological ones, are typically constrained by the number of actions that can be expressed simultaneously, particularly given the bottleneck found in motor space. Take my previous example; the gazelle should either walk towards the new patch of food, or run away in the opposite direction from the predator, critically, it cannot do both. Therefore, behavior requires control algorithms that swiftly and appropriately resolve conflicts between channels requiring incompatible access for the same resources, be them physical (*e.g.,* muscles) or cognitive (*e.g.,* attention and working memory). The problem becomes all the more complex if one considers that incompatibility between channels might not

be absolute. Actions might be partially aligned or even largely independent (*e.g.,* most humans can simultaneously walk and talk on the phone).

We thus reach the core problem that an AS mechanism must solve: *How to arbitrate - i.e., choose - among a set of two or more currently considered actions that might seek access to a common resource.*

1.1.2 Outline of an action selection solution

A successful AS mechanism must be able to solve the AS problem. It should be able to, from a group of multiple potentially incompatible actions seeking access to the same resources, select which ones should be allowed to be expressed. What might a mechanism designed to perform this computation look like? From the gazelle example proposed above, I argue one can intuitively derive some of the features that might be required for an efficient and biologically plausible general AS mechanism:

From first principles, an action selection process should be aiming to select motor programs that achieve a given goal (*e.g.,* consume food or run away from a predator). In order to generate adaptive behavior, such a process of selection must not be left to chance. Returning to the gazelle example, both actions of running away or approaching the patch of food are, in principle, simultaneously valid, albeit achieving frankly different goals: escaping a potential death threat and collecting food, respectively. As opposed to randomly selecting an action, an alternative solution to this problem is to pick whichever channel has the highest *urgency* or *salience*. Under this scenario, both qualities could be associated with the intrinsic motivations of the animal (*i.e.,* hunger levels) but also to the sensory information it has access to from the environment (e.g., how certain of the existence of a predator, or how much greener is the new patch of food). More generally, autonomous adaptive agents should employ an action selection process that ought to be optimal with respect to some goal (*e.g.,* surviving), an assumption often taken when constructing algorithms for artificial agents or modeling the behavior of biological ones (K. Gurney et al., 1998). Animals have indeed been found to behave optimally, a classical example being *matching* behavior

(Herrnstein, 1961), observed in multiple animal species, in which subjects allocate their responses in proportion to the reward rate of each option. However, clear deviations from optimal matching behavior have been similarly found (reviewed in (McDowell, 2005)). These observations of seeming irrational behavior might not stem from a suboptimal decision making process *per se* but instead from a wrongly interpreted objective being optimized or unknown computational constraints of the system, a point I will come back to in the general discussion of this thesis.

Under some scenarios, arbitrating between several hardwired drive→action channels is sufficient to generate strikingly complex behavior (brilliantly demonstrated in (Braitenberg, 1986)). Conversely, one of the most amazing hallmarks of biological agents' behavior is the ability to learn and modify the expected outcomes that result from taking a given action in certain contexts and vice versa(Moore, 2004; Morand-Ferron, 2017). Indeed, while some behaviors can certainly be argued to be innate and somewhat reflexive, most of the action repertoire displayed by animals, in particular Humans, must be learned and updated over a life-time. Notwithstanding, such behavioral plasticity gives rise to a second order problem that must be addressed by an AS mechanism: How should expectations for a given action be learned and updated? Being autonomous, agents can seldom rely on the existence of an *all-knowing* teacher to guide learning. Instead, they must update expectations using feedback from their own past experience. As we will see later, the ability to *reinforce* certain behaviors at the expense of others allows agents to adaptively improve their performance over time, a skill that becomes all the more critical when inhabiting an ever-changing dynamic world where expectations and behavior must be continuously updated (Tierney, 1986).

1.1.3 The need for suppression

All the features stated thus far reflect the *positive* aspect of an AS process in that they describe the necessity to invigorate, or bid for, an action in a sort of positive feedback loop. However, suppression, in addition to promoting, is a

crucial mechanism for action selection (J. J. A. van Iersel & A. C. Angela Bol, 1958; D. J. McFarland, 1969).

An additional key feature of an action selection mechanism is the ability to forbid access to some resources, while relinquishing it to others (Prescott et al., 1999). This is a complex and multi-layered problem. At the most basic level, such ability is critical since simultaneously considered actions are often in conflict with one another (*e.g.,* the classical example of freeze, flight or fight in response to threats). A simple heuristic to solve this dispute is to implement what is known as a *winner-takes-all* computation. Essentially, among N actions, select the one that gathers the most support (perhaps weighting it on the basis of *salience* and agent's *past experience*), regardless of the *value* of all the other alternatives. Crucially, in order to achieve *clean switching* between behaviors this mechanism must be decisive, or else risk non-selected actions interfering with the performance of the selected one. Hence, once a *winner* has been arbitrated, suppression of other non-selected, yet also supported, actions should be guaranteed. Such a mechanism is hard to conceive of without invoking a source of inhibitory control.

At a higher level, on the other hand, one must consider that biological agents rarely use a single resource at a time. Instead, parallel effector systems afford largely independent control (*e.g.,* I can move my two arms largely independently). Likewise, it is increasingly appreciated that the brain is hierarchically organized in that circuits operating at different levels of abstraction interact to generate behavior (Botvinick, 2007; Merel et al., 2019). Therefore an AS mechanism must not only be able to select one action among several competing ones, but instead arbitrate 'compatibility' and 'suitability' within and across this layered infrastructure (Prescott et al., 2007). Finally, it should be noted that this process must not only control low-level mechanical effectors (*e.g.,* muscles), but might also be involved in arbitrating between attentional resources in order to bias action selection to particular goals.

Several authors have also pointed out that the ability to keep some drives under suppressive control might additionally afford actions some temporal continuity (D. McFarland, 1989). Intuitively, given two drives with similar *salience*,

say feeding and drinking, any small decrease in the internal drive (*e.g.,* through consummatory behavior) to perform either motor plan might quickly tip the balance towards the previously non-selected action. This competitive progress might give rise to erratic and suboptimal switching between actions, referred to as *dithering* in robotics literature (Sahota, 1994).

Finally, agents seldom occupy an environment wherein they are afforded infinite time to make decisions. The gazelle does not have endless time to decide whether to forage - the patch might be consumed by other animals or it might ultimately starve to death - or to run away - the predator might quickly pounce in its direction. It is therefore critical that any functionally plausible AS system is selected to meet these ethological demands. In what ways might these constraints shape behavior, and consequently any implementation of such mechanism? On one hand, being quick to react presents an obvious advantage, especially in environments where competition between individuals is present. One strategy to deal with this pressure is to evolve the ability to *plan* or pre-load a motor plan. This priming of action will likely lead to faster reaction times but, at the same time, it also immediately invokes the need for an action suppression system that, instead of inhibiting other competing actions, keeps the currently latent plan from being prematurely released (Hikosaka et al., 2000). Conversely, even though acting fast might pose an obvious evolutionary advantage in some scenarios, others might benefit from delaying the commitment to act altogether. Consider an environment governed by noisy sensory information for instance. In this context it might be adaptive to wait and average information over longer periods of time before committing to a final decision. A suppressive mechanism could, in theory, prevent any action from being unwantedly expressed before the time is deemed right (Cisek et al., 2009; Carland et al., 2019).

1.1.4 Proactive and reactive suppression

Most of us can probably relate to the kid that sits in the classroom holding off the urge to answer the teacher's question, the runner that prematurely leaves the starting block, or the person trying to sneak the last slice of cake just to stop on her tracks once she notices someone staring at her. In general, adaptive

behavior must often rely on inhibition to reduce a tendency to perform a prepotent response that might be considered inappropriate. The study of this ability to withhold inappropriate action is critical since lack thereof is an hallmark of many psychiatric conditions including Huntington's, Tourette's, ADHD (Attention deficit hyperactivity disorder), OCD (Obsessive-Compulsive Disorder) and addiction (Nigg, 2001; Bechara et al., 2006; Chambers et al., 2009). The literature on this topic is vast, hence, for the sake of brevity, I will choose to focus on a single axis that separates inhibition control over action in two large classes: *Reactive* and *Proactive* suppression (or inhibition). I will additionally attempt to provide relevant information on behavior paradigms employed to study these two mechanisms.

Reactive suppression relates to the ability to abort an already initiated action in response to a sensory cue. This type of suppression has been historically associated with a general halting of motor output. Indeed it has been argued that such ability evolved in a context wherein stopping (or *freezing*) in response to an ethologically relevant stimulus might be advantageous to avoid danger(Aron, 2011).

Reactive suppression has been extensively studied using the *Stop-Signal Reaction Time* task (SSRT) (Logan, 1981). Briefly, subjects are asked to react as fast as possible to a cue, by pressing a button for instance. In a variable percentage of trials (classically around 25%), a second cue appears informing the subjects that they should cancel their response in that trial. The *stop* cue is temporally offset from the first cue by a variable delay, which often means that the subject has already initiated a response upon stop-cue delivery, thus sometimes failing to successfully inhibit it. By dynamically varying the time between the two cues, the experimenter can calculate what is usually referred to as the *stop-signal reaction time*. Short delays between the two cues allow the subjects to easily stop their responses, while longer delays lead to many more failures of action suppression since movements are often at their later stages of preparation or execution. Logan & Cowan (1984) found that this behavior can be well described by an independent race model, assuming that two processes - *going* and *stopping* - are triggered by their respective cues. If the *stop* accumulation process

wins, the movement is aborted, else the action will be performed. While reactive suppression has been extensively studied under the SSRT task where no behavior switching is required (*i.e.,* subjects should simply cancel a response and not select a second one), some authors have argued that this form of control might be leveraged by the brain to cleanly switch between adjacent actions (Aron, 2011). Finally, it is unclear whether the stopping signal leveraged for allowing inhibitory control in this task is action specific (*e.g.,* abort the action to press the button) or general (*e.g.,* stop all movements momentarily). However, experiments argue for the latter, as multiple muscles unrelated with the task response are recruited in *stop* trials (Coxon et al., 2007; Aron & Verbruggen, 2008).

In contrast to reactive suppression wherein cancellation occurs after the action process is already in progress, proactive inhibition endows agents with the ability to withhold a prepotent response from being engaged (Meyer & Bucci, 2016). One of the difficulties when studying this form of inhibitory control are the many faces it can take. Suppressive control might be applied to a planned action until it is finally time to release it. Suppressive control might be applied to a tempting action that should not be taken at all. Suppressive control might be able to shift cognitive resources, such as attention, in a way that improves overall behavior performance. Moreover, this form of suppression may be particularly reliant on cognitive processes of learning and memory since it often requires maintaining the representation of a rule or stimulus that is not immediately available to be extracted from the environment (Braver, 2012). It is therefore perhaps not surprising that different behavior tasks continue to be developed and used to study the many aspects of proactive action suppression. Most of these paradigms leverage the difficulty animals have in successfully inhibiting a prepotent planned response. The *anti-saccade* task is a prime example of such a feature (Hallett, 1978).

The *anti-saccade* task relies on the bottom-up, reflexive, orienting response animals tend to exhibit towards a stimulus that appears in their visual field (Munoz & Everling, 2004). Briefly, in this task subjects are instructed to fixate on a motionless target on the screen. After a random delay, a stimulus is presented to one side of the target. The subject is then reinforced for making a

saccade away from the target. When comparing to a *pro-saccade* version wherein subjects are instructed to saccade towards the target, anti-saccade trials exhibit significantly higher error rates (*i.e.,* failure to inhibit the reflexive response) and longer reaction times, suggesting that extra cognitive effort is necessary to perform this latter version of the task (Coe & Munoz, 2017). Interestingly, most errors that occur in the task are quickly followed by short-latency saccades in the opposite direction. This hints at the idea that failure in this task is primarily associated with the inability to suppress the inappropriate *pro* response rather than generating the correct *anti* motor program. There is little doubt that the *anti-saccade* paradigm requires some level of inhibitory control. However, certain features of its design make interpretation of some of the results difficult. Similarly to the SSRT paradigm, it is unclear whether the signal that allows subjects to inhibit a reflexive saccade is action specific or a general *"hold your horses"* process that allows a secondary response (*i.e., the anti saccade*) to be generated. Moreover, since suppression and response generation take place concurrently, it becomes very difficult to tease these two processes apart, especially their neural correlates. Finally, given the obvious contingency between visual stimulus and reflexive behavior, it becomes even more difficult to disentangle signals that are related to the visual stimuli *per se*, or to the to-be generated action.

In a *Go/No-Go* task (G/NG), subjects are instructed to respond as quickly as possible to a *go* cue, typically a visual stimulus. On a small fraction of the trials, a different stimulus instructs the subjects to withhold their response instead. In contrast to the SSRT, G/NG task, subjects are not explicitly instructed to generate a movement on every single trial. However, given that in most of the trials, the default response is to generate a movement, this action gains a sort of prepotency, which must be suppressed. This latent drive can be systematically changed by varying the proportion of *Go* to *No-Go* trials, thus requiring more or less suppressive control from the subject. Interestingly, instrumental performance in this task might also be influenced by what has been referred to as Pavlovian biases, wherein active responses for rewards (*i.e., go*) are more prepotent than active responses to avoid punishment. Suggesting that this paradigm might involve multiple concurrent systems of learning and decision making (Guitart-Masip et

al., 2014). Similarly to the *anti-saccade* paradigm, the G/NG task, in its simplest form, is not able to distinguish between a general proactive signal from a specific action related proactive suppression signal, since the subject is not cued as to which action it should be performing on every single trial. In Chapter 2 I will show how, by leveraging an evolving internal representation of a decision variable, one can probe action-specific suppression signals in the rodent brain, allowing access to study its neural underpinnings.

Despite all three tasks requiring some form of inhibitory control, they differ when such control is required. In the SSRT, inhibition of a response only occurs after the movement has already been instructed and during late stages of movement planning. Conversely, in the G/NG paradigm, suppression is necessary when the No-Go cue is presented. Given the reflexive nature of the response in the *anti-saccade task*, suppression is likely to be present much earlier. While these requirements will place different demands on any neural implementation of inhibitory control, they are nevertheless likely to share, to some extent, a common neural substrate (Meyer & Bucci, 2016). Indeed, previous experiments changing the proportion of "stop" trials in the SSRT suggest that proactive suppression also plays an important role in this task (Schevernels et al., 2015). Consistent with this idea, basal ganglia dysfunction is often associated with impaired performance in all the aforementioned tasks, providing strong evidence that this area is of great importance for the implementation of action suppression in order to generate adaptive behavior (Dillon & Pizzagalli, 2007; Meyer & Bucci, 2016).

1.2 Basal ganglia anatomy

In the 17th century, a British doctor by the name Thomas Willis, examining corpses of patients who died from long paralysis made the astute observation that areas we now take to belong to basal ganglia circuits exhibited clear morphological changes when compared to other patients. Based on his observations, he would later write:

"For the Animal Spirits love to expatiate themselves within these medullous Bodies, (sic.) [Corpora Callosa and Striata], and when they smoothly flow in one series from the two extremes attending the Corpus Callosum (...) towards its middle part, they represent pleasant imaginations and fancies: and when in another series and haply by other Pores, they flow from the midst of the Corpus Calosum into the Gyri of the Brain, they carry thither the signatures of notions for the memory; and when they direct themselves thence into the Corpora striata, and origines of the Nerves, they actuate all the moving parts, and as often as there is occasion, convey to them the Instincts of setting upon motions"

Willis (1685)

This BG-*centric* view of motor control, albeit certainly incomplete - especially given later studies on the role of motor cortex in the control of voluntary movement from Fritsch and Hitzig in late 19th century (Gross, 2007) - would turn out to be certainly prescient.

By the middle of the 20th century a large number of movement disorders were medically documented, including Parkinson's disease, Tourette's syndrome along with different forms of chorea and dystonia (Lanska, 2009). While BG nuclei function was known to be compromised in all these pathological states, little to no neurobiological insight existed to explain the variety of symptoms observed across the different diseases. André Barbeau, a French Canadian neurologist known for his research into Parkinson's disease, when describing the then recent treatment options to some of these BG pathologies, wrote:

"But what of the results? Many are improved that a few years ago would have been miserable, many are permitted a more active life and are forever grateful... but none are cured! The case and exact pathology of the various diseases grouped under the extra pyramidal system remain mysteries almost as deep as in the days of Sydenham and Parkinson. Many clinical varieties have been observed, many pathological studies carried out, but the suffering humanity still goes on twisting, shaking, writhing, jumping and jerking when it does not want to."

Barbeau (1958)

it was not until the 1980s, informed by the aforementioned observations that BG lesions were often associated with movement disorders and the then recent physiological recordings of multiple BG nuclei activity, that the first coherent models of BG anatomy and function were proposed (Penney & Young, 1983; Albin et al., 1989; DeLong, 1990).

Moreover, parallel observations of animals, namely humans, with basal ganglia dysfunction revealed behavior phenotypes that were not merely the result of a movement deficit and instead appear to be compatible with BG playing a larger role in learning, attention, decision-making and other cognitive processes (Loas & Krystkowiak, 2015; Miyashita et al., 1995; Knowlton et al., 1996). Together, these findings seeded the current view that BG, as a functional entity, are a set of subcortical nuclei that regulate the appropriate selection and production of actions depending on expected outcomes (Albin et al., 1989; DeLong, 1990; Doya, 1999).

Classically, the BG circuitry is described as receiving extensive cortical and thalamic input, and in turn exerts its influence on movements via inhibitory innervation to areas involved in movement execution, namely cortex (Albin et al., 1989) (through thalamus) and brainstem motor areas such as the superior colliculus (Hikosaka et al., 2019) and mesencephalic locomotor region (MLR) (Roseberry et al., 2016).

The circuitry is composed of input (**striatum** and subthalamic nucleus, **STN**), output (internal segment of the globus pallidus, **GPi**, and substantia nigra pars reticulata, **SNr**) and multiple intrinsic nuclei (external segment of the globus pallidus, **GPe**, substantia nigra pars compacta, **SNc**, and ventral tegmental area, **VTA**) structures. Taxonomically, it is highly preserved among vertebrates, from mammals, to cyclostomes (lampreys) (Stephenson-Jones et al., 2011), suggesting it was present in the early days of vertebrate evolution (Grillner & Robertson, 2016) and highlighting a putative common solution for the prevalent problem of action selection. Basal ganglia circuit function relies on a disinhibition mechanism. Tonically active output structures of the BG send inhibitory projections to downstream motor regions in the brainstem and thalamus thereby exerting a default suppressive signal thought to keep unwanted motor plans from being engaged. The activity of the output nuclei is, in turn, bidirectionally regulated by two major feedforward basal ganglia pathways: the direct and indirect pathways, which have been shown to promote and inhibit motor function, respectively (Kravitz et al., 2010). In this section, I will review how the structure of the BG is suitable to support the regulation between action production and suppression. I will focus primarily on data from the primate and rodent fields. Although some minor differences are observed across species, both anatomical and physiological characteristics are largely identical. As a result I believe that the gain of including findings from multiple animal models far outweighs the risks of potentially imprecise generalizations. Finally, despite recent anatomical observations that highlight non-canonical projections among BG nuclei (Simonyan, 2019), I will focus, for the most part, on its classical description (Figure 1.1), (Albin et al., 1989; Gerfen, 1992; Mink & Thach, 1993; Y. Smith et al., 1998).

1.2.1 Input structures

1.2.1.1 Striatum

The striatum sits in the forebrain and is the major input structure of the BG, receiving excitatory input from virtually all the cortical mantle (extensively from layer V, but also from other layers), thalamus, hippocampus and several other

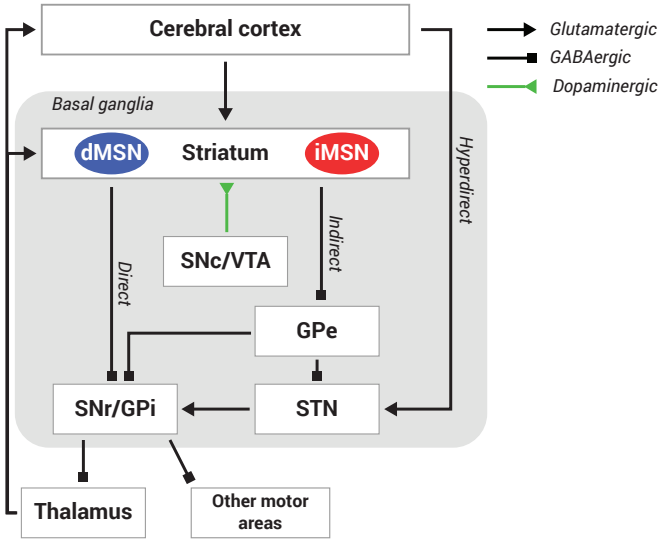


Figure 1.1. Classic "box and arrows" scheme of basal ganglia circuitry. Striatum receives excitatory input primarily from cortex. Two distinct populations of medium spiny neurons in the striatum originate the direct (dMSNs) and indirect pathways (iMSNs). A third hyperdirect pathway results from direct excitatory connections from cortex to STN. The output of the BG, SNr/GPi is opposingly regulated via direct and indirect pathways, that regulate the default inhibition onto downstream motor areas. Plasticity at cortico-striatal synapses is modulated via dopamine input.

limbic areas (Hunnicutt et al., 2016). While the striatum of a given hemisphere receives cortical input from both sides (through pyramidal neurons of the intratelencephalic pathway, IT) a major ipsilateral bias is observed (through both the IT and pyramidal tract, PT, pathways).

Following the general topographic organization of cortical function, striatum is similarly laid out in functionally organized subregions that are determined by the pattern of inputs they receive. Importantly, this topographical organization is maintained, to a large extent, throughout the basal ganglia circuitry nuclei Figure 1.2, giving rise to several parallel computation “channels” within basal ganglia (J. Lee et al., 2020), a feature that I will come back to.

Grossly, striatum is split into two large areas in both rodents and primates: the Caudate-putamen (CPu) and the Nucleus Accumbens (NAc) (Averbeck et al., 2014; Berendse et al., 1992; Selemon & Goldman-Rakic, 1985). While CPu lacks clear anatomical boundaries in rodents, in primates it is further divided in the Caudate Nucleus (Cd) and the Putamen (Put) by white matter fibers that make up the internal capsule. These structures represent the three largest anatomical and functional clusters in striatum. Putamen occupies a dorsolateral portion of the striatum, and receives extensive sensorimotor input from cortex and thalamus, whereas the Cd sits at the dorsomedial part of the striatum and receives unproportionally more input from associative brain areas. Even though it is unclear whether anatomical segregation of CPu serves any functional role, rodent homologous subregions are observed with identical cortico-striatal connectivity referred to as dorsolateral (DLS) and dorsomedial (DMS) striatum, respectively. For simplicity, throughout text I will opt to use DLS/Put and DMS/Cd interchangeably.

A large population of GABAergic medium spiny projection neurons (MSN) make up 90-95 % of all neurons in striatum (Kemp & Powell, 1971; Gerfen et al., 2013) and are thought to be the only cells projecting outside this structure. MSNs exhibit a very negative resting membrane potential (-80mV, (Kreitzer & Malenka, 2008)) driven by inwardly rectifying potassium current (Steiner & Tseng, 2016). It is thought that this inward rectifying current works to shunt weak inputs and instead requires the coincidental activity of multiple inputs to the same neuronal dendritic tree to trigger action potentials. These striatal projection neurons exhibit two distinct efferent patterns that give rise to the two major feedforward pathways in the BG. The first population of MSN (dMSNs) directly inhibits the output structures of the basal ganglia (GPi and SNr) and originate the direct, or striatonigral, pathway. Increases in activity in this pathway decreases inhibitory tonus of the BG output nuclei and consequently increases in motor output . Conversely, indirect, or striatopallidal pathway, medium spiny neurons (iMSN) target the intermediate nucleus GPe before converging on the same output nuclei as the direct pathway. Indirect pathway activity increases the default inhibition

on downstream motor circuits and is thus associated with decreases of movement (Kravitz et al., 2010).

While the two populations are otherwise indistinguishable in terms of cell morphology and, for the most part, spatial distribution within the striatum (Flaherty & Graybiel, 1993), they exhibit marked distinguished neurochemical features. Direct pathway MSNs express the dopamine D1-type receptor (D1R), dynorphin and substance P, whereas striatopallidal MSNs express D2-type receptor (D2R), adenosine A2a receptor (A2a) and enkephalin (Steiner & Tseng, 2016). The expression of dopamine receptors in particular endows the two populations with opposite responses to dopamine release. D1R-expressing dMSNs have their excitability increased by dopamine (Hernández-López et al., 1997; Lahiri & Bevan, 2020) whereas D2R-expressing iMSNs display decreases in activity with higher extracellular dopamine concentrations (Hernandez-Lopez et al., 2000; Nicola & Malenka, 1997; Nicola et al., 2000). In addition to directly influencing excitability of MSNs, dopamine also plays a critical role in regulating spike-time dependent plasticity (STDP) at cortico-striatal synapses (Kreitzer & Malenka, 2008). This is a crucial mechanism thought to support animal adaptive behavior as I will cover later (Shen et al., 2008). Interestingly, the effect of dopamine, similarly to the effect in excitability, is of opposite sign in the medium spiny neurons of the two pathways. Cortico-dMSN synapses undergo long-term potentiation (LTP) upon phasic increases in dopamine, whilst LTP is observed at cortico-iMSN synapses when dopamine falls below baseline levels (Shen et al., 2008; Yagishita et al., 2014; K. N. Gurney et al., 2015; Iino et al., 2020; S. J. Lee et al., 2021). Finally, in addition to endowing the two striatal projection neuron types with distinct neurophysiological properties, the existence of different biochemical markers has also been leveraged to gain genetic cell-type specific access to each of these two populations (Gerfen et al., 2013).

Another, often neglected, organizational feature of the dorsal striatum is its patch-matrix organization. These structures cluster spatially but are only distinguished on the basis of their histochemical characteristics (Brimblecombe & Cragg, 2017). In spite of having been described by Ann Graybiel more than 40 years ago (Graybiel & Ragsdale, 1978) their role in BG normal function is

poorly understood. However, given their distinct pattern of input and output connections, some authors argue for a functional specialization between these two structures (Houk et al., 1994).

The remaining 5-10% of neurons in the striatum are interneurons (Tepper & Bolam, 2004). Novel interneuron subtypes continue to be identified in the striatum, nevertheless two major populations can be identified as cholinergic and GABAergic interneurons. Despite their relatively small number, these cell-types arborize and synapse extensively within the striatum and are able to drastically modulate the activity of the projection neurons (Steiner & Tseng, 2016). While neurophysiological properties and interactions of interneurons with MSNs have long been described, relatively few studies have looked at the activity and causal role of striatal interneurons during behavior, thus making it rather difficult to elucidate on their functional role. Fortunately, recent advances in both recording techniques and genetically defined access to these populations make them attractive targets for future studies (Mallet et al., 2019).

1.2.1.2 STN

The second entry point of the basal ganglia is the STN. As opposed to the inhibitory nature of synapses that govern other BG nuclei connectivity, STN exclusively provides excitatory glutamatergic projections to globus pallidus and SNr. Moreover, few to no interneurons are observed in this structure (Nauta & Cole, 1978). STN exhibits a dramatic reduction in cell number when compared to main BG input, the striatum (1:60 and 1:200 cells in primate and rodents, respectively), yet its importance for normal BG function should not be understated. Pathological STN activity has long been observed in Parkinson's patients, and motor benefits following lesions or deep-brain stimulation targeted to this nucleus remain one of the most successful therapeutic strategies to this day.

Neurons in STN received inhibitory input from GPe, completing the indirect pathway circuit motif before exerting its influence on the output BG structures. In addition, STN also receives extensive cortical excitatory input. The functional role of the so-called "hyper-direct pathway" due to the direct, striatal bypassing,

cortico-subthalamic connections (Nambu et al., 2002) that give rise to it, is still very much debatable. Technically, due to the reentrant nature of the circuit, it has been somewhat difficult to tease apart the specific contribution of indirect pathway activity from the monosynaptic cortical input (hyper-direct pathway) to STN function, and eventual influence on its outputs. Two relevant features of this cortico-subthalamic pathway might inform its function: First, due to its monosynaptic connectivity, cortical activation leads to very short latency activation of STN neurons and concomitant excitation of BG output nuclei (Polyakova et al., 2020). Second, while still following the general topological organization of BG, STN neurons arborize relatively broadly into GPi/SNr, providing a source of inhibition to downstream motor targets. It is therefore not surprising that multiple studies have now implicated the hyper-direct pathway on the ability of animals to reactively and quickly stop in response to sudden cues supporting reactive suppression (W. Chen et al., 2020; Mallet et al., 2016; Schmidt et al., 2013; van Wouwe et al., 2020). Despite the obvious behavioral advantages of being able to quickly abort ongoing actions in response to salient external stimuli, this form of inhibition has been shown to be largely non-specific in the sense that it provides a general stopping mechanism rather than a selective one.

1.2.2 Output structures

1.2.2.1 GPi and SNr

The output role of the basal ganglia is shared between two different areas: the internal segment of the globus pallidus (GPi[1]) and the substantia nigra pars reticulata (SNr). Despite their distinct anatomical and developmental origins, these two areas are often considered a "continuous" output structure, since they appear to be functionally, biochemically, and physiologically largely identical. Therefore, I will choose to group them in single, continuous, identity: GPi/SNr (Lanciego et al., 2012).

[1]It should be noted that the GPi homologous structure in rodents is the entopeduncular nucleus, EP (van der Kooy & Carter, 1981). For simplicity, throughout the dissertation, these will be used interchangeably.

As previously mentioned, the output structures of the basal ganglia display a relatively high level of tonic activity (60-80 spikes/s (Steiner & Tseng, 2016)). GABAergic projections from GPi/SNr provide a default inhibitory tonus onto downstream areas of the thalamus, midbrain and brainstem thought to prevent unwanted plans to be deployed. Recordings during awake behavior indeed confirm the modulation of SNr units aligned to spontaneous locomotion. Strikingly, while several neurons are indeed inhibited, the majority is excited (Gulley et al., 1999), potentially highlighting the need to balance action invigoration and suppression during spontaneous behavior.

Activity in these structures is mainly regulated by the relative contribution of direct and indirect pathways, through direct striatal GABAergic and STN excitatory/GPe inhibitory input, respectively. Interestingly, while optogenetic activation of medium spiny neurons of the two pathways produces changes in firing rate consistent with the classical BG model, cases of unexpected modulation are observed. These are often delayed responses and likely reflect higher-order synaptic motifs but nevertheless highlight the complexity of the circuit. (Deniau & Chevalier, 1985; Kravitz et al., 2010; Freeze et al., 2013).

1.2.3 Intrinsic nuclei

1.2.3.1 GPe

Indirect pathway medium spiny neurons do not project directly to the output nuclei of the basal ganglia, GPi/SNr, and instead, synapse onto intermediate structures. The major recipient of these inhibitory projections is the external segment of the globus pallidus (GPe) (Nóbrega-Pereira et al., 2010; Hernández et al., 2015). In addition to the striatal inhibitory input, GPe is excited by the STN and, to a much lesser extent, the parafascicular nucleus of the thalamus, cortex and other basal ganglia nuclei (Fink-Jensen & Mikkelsen, 1991; Parent & Hazrati, 1995; Kita & Kitai, 1994; Hazrati et al., 1990).

Studies using immunoreactivity and electron microscopy have shown that, save for a small fraction of CHAT+ interneurons, all neuronal cells in GPe are inhibitory (Hegeman et al., 2016). Two major GABAergic cell types are found

within GPe. PV-positive prototypic GABAergic neurons make up the canonical feedforward motif by inhibiting STN and occupy the vast majority of GPe territory. These cells exhibit high and stable firing statistics. Arkypallidal neurons, on the other hand, send feedback projections to the striatum and exhibit relatively lower and irregular firing rates. (Mallet et al., 2012; Abdi et al., 2015; Dodson et al., 2015).

The canonical description of the BG circuit depicts GPe as a mere relay station for the indirect pathway that inverts the effect iMSNs have on the output nuclei of the BG. As a first approximation, this model has helped the field with the interpretation of BG function quite well, although relatively more recent findings of non-classical direct output to GPi/SNr, or pallidostriatal feedback projections, might challenge the functional reductionist view of the circuit (Sato et al., 2000; Kita et al., 1999; Kita & Kitai, 1994; Mallet et al., 2012).

1.2.3.2 VTA and SNc

The main source of feedback projections within the basal ganglia are the Ventral Tegmental Area (VTA) and Substantia Nigra pars compacta (SNc). Both areas project densely to virtually all areas of the striatum where they release dopamine. Similarly to all other areas within the BG, VTA/SNc also maintain a rough topographic organization albeit less strict given dopaminergic neurons' extensively broad arborization (Joel & Weiner, 2000). Specifically, VTA projects preferentially to NAc while SNc projects to CPu. This latter projection is also topographically organized in that lateral areas of SNc project to dorso-lateral-posterior areas of the dorsal striatum (Beckstead et al., 1979). This organization has been somewhat debated however, and other authors propose a different organization where a strong communication across BG loops is observed (Haber et al., 2000; Haber, 2003)

DAergic neurons have been implicated in a wide variety of functions ranging from motor control (Carlsson et al., 1957), motivation and reward seeking behavior (Mogenson et al., 1980) and, critically, learning (Schultz et al., 1997). In particular, neurons in both areas have been shown to phasically respond to

unpredicted rewards, or reward predicting cues, computing a quantity known in the reinforcement learning literature as "reward-prediction error", the difference between experienced and expected reward (Montague et al., 1996; Schultz et al., 1997). This finding has not only been observed in DAergic cell bodies but has also been replicated when directly measuring dopamine release at striatal terminals (Day et al., 2007; Gan et al., 2010). Changes of dopamine concentration in striatum support one of the pivotal roles attributed to BG - *learning* - where it has been shown to be able to modify glutamatergic synaptic contacts between medium spiny neurons and its inputs (J. N. Reynolds et al., 2001; J. N. J. Reynolds & Wickens, 2002; Shen et al., 2008).

Finally, in addition to DAergic neurons, a second large population (30% of all neurons) of GABAergic interneurons can also be found in both structures and are thought to be critical for the computation of temporal-difference error (Cox & Witten, 2019).

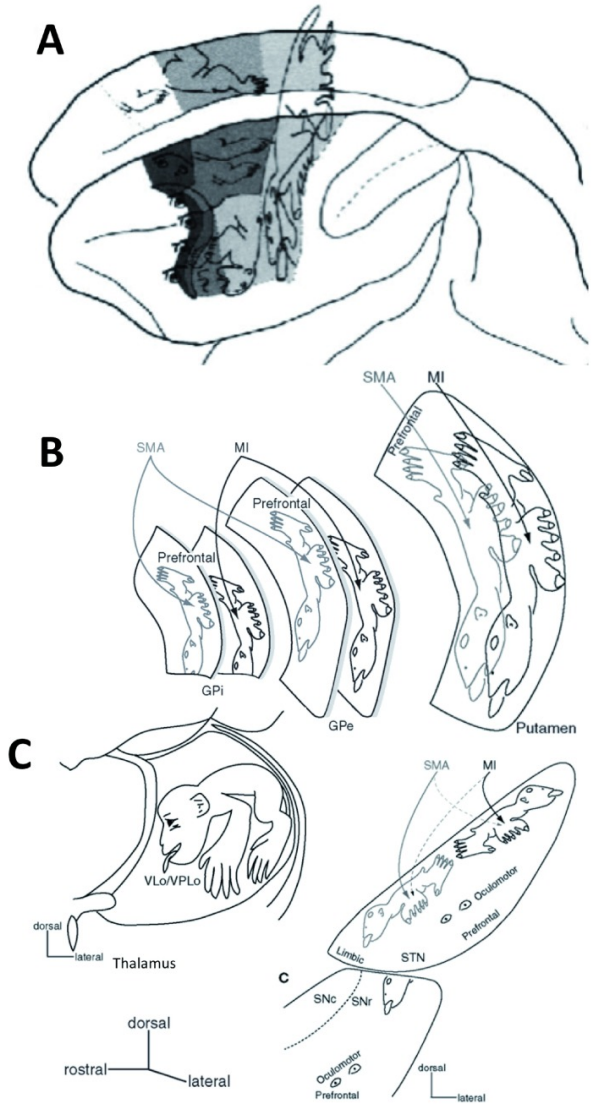


Figure 1.2. **a)** Somatopic representation of body regions in monkey cortex. **b-c)** Somatotopic body representations in vaious BG nuclei in the sensorimotor loop. Reproduced from Simonyan (2019).

1.3 Basal ganglia organizing principles

With the introduction of the key structures that compose the basal ganglia, I will now attempt to summarize what are, in my view, the anatomical and physiological features of BG circuitry that support its functional role in motor function, decision making, reinforcement learning, and cognition.

1.3.1 Recurrent parallel circuits

BG circuitry is organized in a large(-ly) recurrent circuit. Input from cortex, thought to carry contextual information about the state of the world and available actions, excites the large population of GABAergic neurons in the striatum which, in turn, bidirectionally influence the activity of basal ganglia output through the direct and indirect pathways. Part of this "open-loop" output will project to low-motor areas where it is thought to regulate parameters of movement execution. However, a significant portion of SNr/GPi output is reentrant. This loop is eventually closed through inhibitory SNr/GPi projections to thalamus and a subsequent excitatory thalamo-cortical synapse. Critically, activity within each loop is plastic. Namely, cortico-striatal synapses can be potentiated/depotentiated following actions or events that resulted in better or worse than expected outcomes, a teaching signal shown to be relayed by midbrain dopaminergic neurons (Montague et al., 1996). The ability to modify these synapses affords animals the ability to modify their future actions as a function of past experience, a linchpin for adaptive behavior.

This general recurrent motif seems to be repeated across parallel channels in the BG, suggesting a similar computation performed on top of different inputs. Some debate exists as to whether these loops are convergent or segregated. These two extreme views differ in the degree with which information is partitioned and potentially integrated among the multiple functional domains of the BG. Proponents of functional convergence hold that the lateral connectivity found in the striatum, together with the overall anatomical convergence along the main BG axis, would support integration of information among distinct functional domains (Bolam et al., 1993; Percheron et al., 1994). However, while lateral connectivity

was indeed found in striatum (Taverna et al., 2008; Planert et al., 2010; Klaus et al., 2017) and in the palladium (Percheron et al., 1993), and extensively used to argue for models of "winner-takes-all" action selection, such inhibition was found to be weak or non-existent (Jaeger et al., 1994). Additionally, an overwhelming amount of anatomical and physiological evidence support the idea that these loops are largely segregated and not convergent. Projections from striatum to SNr and GPi generally uphold the topographic principle found in sensorimotor cortico-striatal projections (Deniau et al., 1996; Alexander et al., 1986; Romanelli et al., 2005; J. Lee et al., 2020) [2], and even in the rare cases where some degree of anatomical convergence across distinct domains was found, it seemed to be circumscribed to functionally related areas(Foster et al., 2021). Finally, Alexander & Crutcher (1990) have also found that, within each loop, neuronal responses are functionally related (*e.g.,* response to leg movements).

Together, while it is likely that a certain degree of integration across distinct areas will be found throughout the main axis of the BG, the data suggests that, at least in the sensorimotor loops where a somatotopic map is maintained throughout BG nuclei, information is carried out in parallel functional sub-domains, raising the important question on how might information across distinct functional loops be combined to guide behavior. A possibility is that sharing of information across domains is implemented at the level of intracortical communication (*e.g.,* prefrontal cortical areas are known to project to premotor areas (Lu et al., 1994)) or through previously described divergent projections from thalamus to different cortical areas (Rausell et al., 1998; Morel et al., 2005; Hunnicutt et al., 2014).

Originally, three large loops were suggested to exist, based on the three large input areas in the primate striatum. These were called "motor", "association" and "limbic" loops, supported by the Putamen, Caudate, and Accumbens, respectively. Not much time passed until more functionally specialized loops were suggested, namely the oculomotor, (Hikosaka et al., 2000) involved in guiding saccadic behavior. If anything, recent studies highlight the finer and finer struc-

[2] This organization principle is somewhat debatable. Projections to GPi do maintain a strong topological organization. However, some authors have argued that in SNr, this principle is somewhat relaxed and more overlap between different, yet functionally related, areas is found (Joel & Weiner, 1994; Foster et al., 2021).

ture of BG functional organization (Hamid et al., 2021; Menegas et al., 2018; Hunnicutt et al., 2016; Foster et al., 2021).

Finally, at a meta-scientific level, in light of the frankly similar circuit architecture shared across BG sub-circuits, it seems highly plausible that different sub-domains perform comparable computations. As a result, despite the different input information accessible to each channel - thus explaining the heterogeneous activity seen across loops -, the algorithmic understanding of the computations carried out in one of these channels will likely provide critical insight on the computations performed in other parallel circuits of the BG.

1.3.2 Anatomical convergence

A second key, related, feature of BG circuitry is its anatomical convergence along the main circuit axis. In each step of the pathway, a large drop in the number of neurons is found. Concretely, the number of cortical neurons projecting to the striatum was found to be two orders of magnitude greater than the striatal neurons (Kincaid et al., 1998), which in turn are two orders of magnitude greater than GPi neuron counts (Percheron et al., 1993), which is thought to be *de facto* "bottle-neck" within BG circuitry. It is important to reiterate that, given evidence provided in the previous section, such funneling is likely the result of dropping the number of neurons within a given functional domain and does not seem to support the view of functional integration between distinct striatal areas (Kemp & Powell, 1970; Yeterian & Pandya, 1991). Consistent with this idea, the overlap in the striatal projection zones from nearby cortical regions was found to decay exponentially with cortical distance (Averbeck et al., 2014). Similarly, activity correlation between pairs of cells in the striatum have been found to be relatively weak (Jaeger et al., 1994; Bar-Gad, Heimer, et al., 2003) and to decrease with a very steep spatial constant (Klaus et al., 2017). A possible reason for the lack of activity correlation between neighboring neurons is a sparse input from cortical areas that endow striatal neurons with different activity profiles (Kincaid et al., 1998). With this in mind, several studies have proposed that BG is effectively performing some form of dimensionality reduction on its inputs (Bar-Gad, Morris, & Bergman, 2003). This progressive information decorrela-

tion, similarly to what has been modeled in other brain areas (Maltenfort et al., 1998), is thought to achieve efficient information compression. Moreover, by having access to a form of teaching signal that is thought to support adaptive behavior (*i.e.,* reward prediction error), some have suggested a mechanism wherein compression is biased towards cortical inputs that are in some way relevant to the current behavior repertoire of the animal, effectively implementing a "reinforcement-driven dimensionality reduction" algorithm (Bar-Gad, Morris, & Bergman, 2003; Joel et al., 2002). The role of this putative dimensionality reduction is still very much debatable, however, given anatomical and energetic constraints principles argued to guide nervous system development and evolution (Sterling & Laughlin, 2017) along with relative low-dimensional array of effectors animals have at their disposal, it might be advantageous to bias action-selection in a relatively lower-dimensional space enriched in relevant features and invariant to others (Motiwala et al., 2020). Additionally, invariance to some features of the world and actions might afford animals some degree of generalization which might accelerate learning in some scenarios (Bar-Gad, Morris, & Bergman, 2003). For example, the action to eat a pear or an apple does not mechanically differ substantially, moreover both fulfill a similar goal of reducing hunger level. It is thus perhaps advantageous to bundle both percepts to the same downstream motor command.

1.3.3 Disinhibition

GPi/SNr GABAergic neurons are known to exhibit very high firing rates at rest (40-80Hz, (Mink & Thach, 1991; Wichmann et al., 1999)). This has led to the idea that BG keeps downstream thalamus, midbrain and brainstem areas under a default, tonic, inhibitory tonus. Seminal studies have since shown that selectively inhibiting SNr areas, known to project to the superior colliculus, is sufficient to increase the likelihood of saccadic movements (Hikosaka & Wurtz, 1985). Similarly, in a non-human primate model of Parkinson's disease, a BG pathology primarily characterized by akinesia, focal application of the GABA(A) agonist muscimol in GPi has been reported to ameliorate motor symptoms (Baron et al., 2002). These findings paint a picture wherein BG *modus operandi* is primarily through disinhibition, a mechanism through which it is thought to regulate the

appropriate selection of action (Mink, 1996; K. Gurney et al., 1998; Grillner et al., 2005).

The level of disinhibition is oppositely modulated by the direct and indirect pathways of BG (Freeze et al., 2013). Direct pathway medium spiny neurons' activity has been shown to induce brief pauses in GPi/SNr via which actions are thought to be selected. Conversely, indirect pathway activity leads to short-latency increases in the discharge rate of GPi/SNr cells, and is thought to raise the threshold for action. Such functional opponency has not only been observed with regards to overall motor output but also with respect to ongoing motor sequence production (Sippy et al., 2015; Tecuapetla et al., 2016), reinforcement (Kravitz et al., 2012; Yttri & Dudman, 2016), and value-based decision making (Tai et al., 2012). However, in what has been used to argue against this functional opponency view, coactivation of dMSNs and iMSNs is observed during action production (Tecuapetla et al., 2014, 2016; Cox & Witten, 2019). Interestingly, long before these observations, Jonathan Mink theorized that action production would proceed with the simultaneous and concerted activation of direct pathway, responsible for invigorating actions to be taken, and indirect pathway, that would inhibit other, potentially competing, actions (Mink, 1996) .

Such a mechanism would, given the relatively more focused inhibition from Striatum to GPi when compared to the excitation from STN, give rise to a "center-surround"-like filter, reminiscent of the ones found in early visual system (Kuffler, 1953). The functional consequence of this wiring motif is still very much debated, however, one could imagine how such computation would support the selection of a given motor plan over others by selectively increasing the signal-to-noise ratio between actions. The generalization of this hypothesis has led to the idea that similar selection of cortical activity might be happening in other parallel loops of the BG essentially regulating the activity of more abstract, cognitive and limbic cortical states. Finally, "*selection through disinhibition*" is one of the main arguments for BG selecting, as opposed to generating, actions. In this view, BG would perform a computation akin to *vetoing* motor plans, that are currently being considered somewhere else in the brain circuitry, and allow/prevent their expression (Mink, 1996; Doya, 1999; Hikosaka et al., 2000).

1.4 Basal ganglia & action selection

There is certainly no prescribed need for a centralized action selection mechanism in the vertebrate brain. Indeed, in many cases it can easily be argued that a distributed decision making and action selection process is taking place (Cisek, 2007; Steinmetz et al., 2019).This issue becomes all the more complex (and interesting) when one considers that distinct functional brain modules will play a different role to behavior that will, to some extent, contribute to specify parameters of action. Nevertheless, while several areas can ultimately influence behavior (*e.g.,* risking *ad absurdum*, a mammal lacking a retina will not be able to use visual input to guide behavior), it is clear that some areas will have a role more akin to what one would consider an action selection module. In particular, many authors have argued that BG is organized to select desired actions and to inhibit potentially competing ones. Here's why:

Basal ganglia is a hub Any action selection module worthy of its name must be able to optimally select actions. In order to do so, it must be *aware* of the current context the agent is inserted in. The striatum, in particular, receives input from all over the cortex, thalamus and various limbic areas (Hunnicutt et al., 2016; Foster et al., 2021). It is therefore thought to have access to sensory information, currently considered motor plans, models of the world and goals that can be leveraged to guide the appropriate decision of *what to do next*.

Basal ganglia can regulate action production and suppression As previously mentioned, action selection will decidedly benefit from the ability to bidirectionally control the bidding of each competing action. This ability will afford the system to cleanly switch between behaviors, prime actions (*planing*), modulate decision-making integration times and potentially cancel already ongoing movements. One of the most striking features of BG organization is apparent functional mapping between action production and suppression to the direct and indirect pathways, respectively. While many details - physiological, mechanistic and conceptual - of this duality remain to be fully elucidated, it seems safe to assume that, being such a

conserved feature across vertebrates, it indeed represents a useful general mechanism to endow an action selection module with.

Basal ganglia as a centralized action selection module If context information is present throughout the brain what is the advantage of a centralized action selection module as opposed to a distributed one? Two main arguments have been put forward. The first is concerned with energetic and spatial efficiency and it has been thoroughly covered elsewhere (Redgrave et al., 1999). Briefly, each axonal connection between two neurons comes with its own energetic and spatial overhead. Such evolutionary pressure might have selected brains that efficiently wire neural networks. At the extreme, a centralized AS module with N neurons exhibits $2N$ connections (to and from the central selection module). Conversely, a fully recurrent network, shown to be able to perform action selection, grows with N^2 (each unit must connect to and from all others). While networks in the brain have been shown to display a significant degree of local recurrence, for long distances this mechanism is unfavorable. Instead, information from functionally distinct areas could be efficiently funneled through an action centralized module thereby requiring much less *wire*. The second argument is perhaps more subtle. Taking into account the functional specialization observed across distinct brain areas, it stands to reason that different information *channels* might require their own action selection loops to be resolved. This solution might afford modularity in that it might allow partially independent control and learning across the system. Finally from an evolutionary standpoint, it seems that duplicating a motif that already exists (*e.g.,* cortico-BG sensorimotor loop to an associative loop) is a more plausible event than to stumble upon a *de novo* solution (Wagner & Altenberg, 1996; Chakraborty & Jarvis, 2015). A centralized AS module shows promise in answering both demands.

Basal ganglia supports learning That animals can learn seems to go without saying. It is the core skill that affords agents the ability to dynamically adapt to changes in their environment and to ultimately behave adaptively. Any action selection module that hopes to select the best action at any

given time must therefore be able to change preferences towards a given plan when the reward landscape changes. This *policy* (*i.e.* what action to perform for a given state of the world) can be updated in multiple ways, as I will cover in a later section. Conceptually however, in order to *learn*, a teaching signal is required. Unfortunately, autonomous agents seldom have access to a *teacher* that provides labeled examples on how to behave. As a result, animals must instead rely on their own experience to update policies. Intuitively, one form this error could take is the difference between what an animal expected and what it experienced. Positive and negative deviations in this quantity should increase and decrease, respectively, the probability to perform the same action in the future. Dopaminergic neurons in the midbrain encode a quantity that closely resembles this *reward prediction error* and extensively innervate striatum, where dopamine release has been shown to be able to induce plasticity at cortico-striatal synapses, providing an obvious substrate and mechanism necessary for learning.

1.5 Models of basal ganglia

How to determine the function of a specific brain area? Deciphering the computations performed by specific brain circuits is simultaneously one of the most difficult objectives to achieve in systems neuroscience yet one of its ultimate goals. Basal ganglia circuit function in particular, has been one of the most modeled circuits in the mammalian brain. BG models have built on top of anatomical, neurochemical and physiological experimental data to provide a general description of circuit function. Moreover, albeit challenging, and sometimes even perilous, studying disease states to infer circuit function may provide key insight. BG research has found some success with this approach. Several common neurological disorders, especially those with a profound motor effect, have been associated with a dysfunctional BG circuitry.

It will become clear that modeling BG circuit function is a challenging endeavor given the breadth of functional roles attributed to this circuit. Nevertheless, every model, be it computational or otherwise conceptual, is a source

of influence that paves the way to new hypothesis and experiments , forcing the reinterpretation of its compatibility with current the model instantiation and its potential reiteration. Risking the *cliché*: "All models are wrong, but some are useful"(Box, 1976).

This section serves to present the reader with a condensed view over influential model instances that found, to some degree, success in explaining BG function. It is worth noting that while these models take different approaches and sets of assumptions, most of their features are not necessarily incompatible with one another and instead represent somewhat distinct facets of BG function. In fact, in order to arrive at a satisfying mechanistic understanding of BG role I believe a lot is to be gained from combining knowledge from different models.

Finally, far from exhaustive, this section will necessarily be a biased exposition, from both a historical and overall content perspective, but one which I hope will justify my current vision of the BG circuit function discussed later.

1.5.1 Rate model

Originally proposed in the 1980s, the rate model (RM) was one of the first conceptual models that tried to describe general BG function considering what was then known of connectivity, neurochemical and physiological data (Penney & Young, 1983; Albin et al., 1989; DeLong, 1990).

The RM was built on top of key observations and assumptions. First, an intact basal ganglia is necessary for executing or otherwise maintaining motor commands. The providence of these commands is still very much debatable, but cortex was considered a probable source that could be regulated through positive thalamic feedback, disinhibited, in turn, by BG output structures. Second, BG circuitry exhibits two, largely feedforward parallel circuits - the direct and indirect pathways - that bidirectionally regulate the activity of GPi/SNr. Third, dopamine bidirectionally regulates the activity of these two pathways by binding to distinct dopamine receptors (D1R/D2R found in dMSNs/iMSN, leads to neuronal excitation/inhibition, respectively). Fourth, actions are selected by the

focal disinhibition of a subset of motor plans in cortex, or downstream brainstem motor areas, by the regulation of GPi/SNr default suppressive activity.

Initial instantiations of RM proposed that the direct and indirect pathways provide a global motor facilitation and suppressive signal, respectively (Albin et al., 1989; DeLong, 1990). Indeed, early optogenetics experiments have shown that anatomically broad activation of these pathways leads to an overall increase/decrease of motor output in the open-field, respectively(Kravitz et al., 2010). However, the sparseness with which these cells synapse to GPi/SNr led to the idea that activity of MSNs regulates the selection of specific actions rather than overall motor activity *per se*. In other words, at the limit, for each dMSN selecting a given action a same *anti-neuron* iMSN was available to potentially inhibit the same action. Selection would then proceed by regulating the balance between the activity of dMSN and iMSN for the same action (Schroll & Hamker, 2013). Despite its simplistic nature, RM found great success in explaining and predicting observations from several BG related pathologies.

If the balance between direct and indirect pathway activity dictates the expression of action, one would expect pronounced changes in action production behavior in conditions wherein this balance is compromised. Several BG pathologies are characterized by a clear unbalance in the activity of these two pathways, providing testing grounds for this model.

Huntington's disease (HD) is a rare genetic disorder wherein the *HTT* gene suffers repeated trinucleotide repeat expansion, giving rise to a pathological form of the Huntingtin protein (MacDonald et al., 1993). Patients suffering from this disorder exhibit strong motor deficits in the form of uncontrollable jerks or writhing movements (Roos, 2010). Through mechanisms yet to be fully elucidated, several studies have identified an early preferential loss of striatal-pallidal (iMSNs) medium spiny neurons in HD patients (Reiner et al., 1988; Sapp et al., 1995; Deng et al., 2004), thought to lead to an overall decrease in indirect pathway activity. Indeed, electrophysiology recordings from HD patients undergoing deep-brain stimulation implantation surgery have shown significant increases in overall GPe activity and concomitant decreases in BG output structures activ-

ity (GPi) (Cubo et al., 2000; Starr et al., 2008). Similarly, patients suffering from dystonia, a condition characterized by the involuntary, often repetitive, co-contraction of agonist and antagonist muscle groups, low GPi activity has also been described (Starr et al., 2005; Vitek, 2002). Finally, patients ailed with hemiballism, a disorder characterized by violent and large amplitude involuntary unilateral movements, lesions in STN are often documented (Martin & Alcock, 1934; Hawley & Weiner, 2012). In other words, pathologies characterized by the lack of suppressive motor control, otherwise known as hyperkinetic movement disorders, are very often accompanied by a deficit in indirect pathway function. Consistent with the rate model, decreasing this *no-go* signal might be expected to lower the threshold for action selection, resulting in actions being allowed to "slip through the cracks" and ultimately generating an unwanted motor command. Interestingly, some of these conditions are also associated with cognitive deficits that can be described as lack of suppressive control or impulsivity. Together with the fact that other conditions, known to exhibit a similar lack of inhibitory control, specifically attention deficit hyperactivity disorder (ADHD, (Barkley, 1997; Qiu et al., 2009)) and obsessive-compulsive disorder (OCD, (Saxena & Rauch, 2000)) and, similarly, a compromised BG circuit, it suggests that the same computation that is applied to suppress motor output might similarly be able to regulate higher-order cognitive processes such as attention and memory.

RM also predicts that increasing the ratio of indirect-to-direct pathway activity should produce motor deficits consistent with decreases in motor output (*i.e.* akinesia/bradykinesia). In Parkinson's disease (PD) midbrain dopamine neurons, especially in SNc that project extensively to sensorimotor portions of the striatum, suffer selective and progressive degeneration. Consistent with the opposing effect dopamine has been shown to have on the excitability of these two populations, increases in activity of the indirect pathway and parallel decreases in the activity of the direct pathway medium spiny neurons are observed in animal models of PD (Mallet et al., 2006; Parker et al., 2018). Additionally, several studies have found increased firing rates in STN and GPi, consistent with increased/decreased activity of iMSN/dMSNs in Human PD patients (Benazzouz et al., 2002; Hutchison et al., 1998). A corollary is that returning dopamine

concentration to normal levels should restitute this balance and alleviate motor symptoms. Indeed, dopamine replacement therapy is, to date, one of the most successful forms of symptom management in PD patients. At the physiological level, intra-operative administration of L-DOPA reduces GPi firing rate, and doses associated with L-DOPA induced dyskinesia have been shown to produce further decreases (Lozano et al., 2000; Levy, Dostrovsky, et al., 2001). Finally, local inactivation of STN using muscimol restores firing rates to a healthy range and ameliorates parkinsonian symptoms (Levy, Lang, et al., 2001).

Together, these, and numerous other consistent observations, account for a scenario wherein relative levels of activity of the two pathways determine the extent of motor program activation. Direct pathway medium spiny neurons, by briefly silencing GPi/SNr output structures, dishibit thalamocortical or brainstem motor areas increasing the likelihood of movement, whilst indirect pathway activity is associated with increases in the default suppressive signal. While this functional opposition view has been a staple of most models of the BG, some experimental results are not immediately compatible with the simplest version of the rate model. For example, a dysfunctional GPi is associated with motor symptoms, yet pallidotomy in non-human primates is not associated with obvious overt motor deficits in healthy animals (Desmurget & Turner, 2008). This appears consistent with other studies showing that while changes in GPi firing rate are routinely found in these pathologies, these are relatively small. Instead, changes in temporal firing patterns (i.e. bursting and LFP) are more salient and prevalent (Litvak et al., 2011; Eusebio et al., 2009). Finally, in what has been used as the strongest argument against the rate model and the functional opponency view of the two BG pathways, iMSNs and dMSNs were shown to be co-active around movement (Cui et al., 2013; Tecuapetla et al., 2014; Markowitz et al., 2018), revealing how imbalances between the overall activity of the two pathways are not a requisite to generate movement and, instead, co-activation is necessary to perform action (Tecuapetla et al., 2014, 2016). To accommodate these findings new models highlighting the potentially concerted and synergistic activation of the two pathways have been put forward.

1.5.2 Center-surround model

As noted above, the two pathways of the BG exert opposing influence on the output of the BG. Yet, during periods of overt motor output, where one would expect to observe anti- or uncorrelation of the two pathways to facilitate movement, co-activity is instead observed. Incidentally, two decades before the co-activation of the two pathways had been directly observed, J. Mink proposed that action selection should take place with the concurrent activation and suppression of motor programs (Mink, 1996; Mink & Thach, 1993). These two functional processes were, not surprisingly, ascribed to the activity of dMSNs and iMSNs, respectively.

How might these two pathways interact to regulate action selection? Taking into account anatomical and physiological evidence, it has been proposed that BG implements action selection through a center-surround filter, akin to those found in early sensory visual system (Kuffler, 1953), wherein focused striatal-pallidal activity activates a "central" action, while excitation from subthalamic nucleus to SNr/GPi raises the inhibition onto surrounding, perhaps competing actions. Under such a model one would indeed predict co-activation of the two pathways during movement.

Several pieces of evidence support the model. While classical experiments have shown that either pathway originating from the same striatal region converge onto functionally identical domains and the same neurons in GPi/SNr (Hazrati & Parent, 1992b; Bolam & Smith, 1992; Foster et al., 2021), some differences pertaining to the size of their axon arborization have been found (Hazrati & Parent, 1992a). On one hand, single striatal-pallidal neurons show convergent and sparse projections and are thus hypothesized to carry action-specific information. Conversely, subthalamic projections to BG output nuclei are denser and relatively more diffuse. In other words, while general spatial convergence is found, inhibition of motor programs is thought to be broader and more diffuse while excitation appears relatively more focused. Depending on the exact details of this interaction, the outcome will be a "Mexican-hat"-like filter in anatomical and thus, perhaps, in action space (Figure 1.3, (Mink, 1996)). A corollary to this hypothesis is that during action performance both activation and suppression of

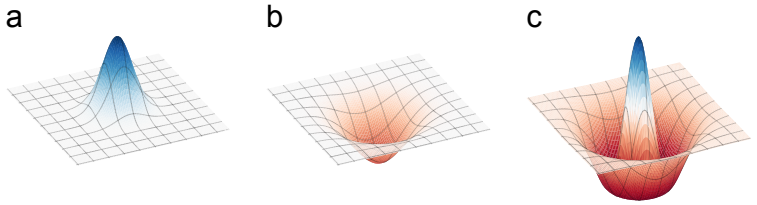


Figure 1.3. Center-surround filter in action space. Blue and red depict activation and suppression of motor programmes, respectively. a) and b) represent the potential influence direct and indirect pathway have on action selection, respectively. c) depicts outcome from the sum of the two filters. Notice the increased sharpness of the center peak that results from this operation.

BG output units should be observed. Indeed, while recording from mammalian GPi/SNr, both responses have been simultaneously recorded (Gulley et al., 1999), suggesting that dishibition alone is not sufficient for movement execution.

The computational role of this anatomical motif is very much up for debate. One idea is that such a filter might increase the overall signal-to-noise ratio between what *to-do* against a noisy background of many potential actions and thus potentially implementing a positive feedback loop with cortex that eventually results in the selection of a single, focused, motor plan.

Despite what appears to be an elegant theory to explain the simultaneous suppression and excitation of BG output units, several questions remain unanswered with regards to the exact implementation of this proposed computation. First, in its simplest form, the theory suggests that competing actions are represented in close anatomical proximity, thereby leveraging the center-surround filter that is also defined in the same anatomical space. This idea has some merit in that, given the overall topographical organization of the circuit, actions requiring access to the same effector (e.g., arm) should also be the actions that inherently impose the highest level of conflict and thus in potential greater need of suppression. This theory is, to the best of my knowledge, untested. Second, recent evidence suggests that, at the level of striatal sub-domains, iMSN-GPe (indirect pathway) projections are actually more convergent than dMSN-GPi/SNr (Foster et al., 2021). Third, while indirect pathway activation does drive the excitation

of GPi/SNr through projections from STN, it is still up to debate whether the broad functionally suppressive signal results from the indirect activation of STN units through the corticostriatal pathway activity (*i.e.* indirect pathway) or from the direct activation of STN units through cortico-subthalamic projections (*i.e.* hyperdirect pathway). Behavior and physiological data support that the so-called hyperdirect pathway implements a general, short-latency (Nambu et al., 2000), halting mechanism that would undoubtedly take advantage of broad suppression of most action programs (Nambu et al., 2002; Schmidt et al., 2013; Dunovan et al., 2015; W. Chen et al., 2020; van Wouwe et al., 2020). Indeed, following intracortical electrical stimulation a triphasic response is often invoked in SNr/GPi. A first short-latency volley of excitation (10ms) is thought to be mediated by a hyperdirect pathway, followed by a slower period of suppression and excitation, thought to be mediated by direct and indirect pathways, respectively (Nambu et al., 2000; Kita & Kita, 2011).

Finally, it remains to be determined whether this motif represents a hard-wired mechanism that meets the underlying biophysical constraints of the circuit necessary for action selection, or to what extent it can be modulated through learning. Nevertheless, given the overwhelming amount of evidence supporting the role of BG in not only selecting what actions to perform but also being necessary to *learn* what actions to produce, the activity of these BG circuits are very likely plastic, allowing dynamic changes contingent on the animal's previous experience, thus guiding future behavior.

The next section will focus precisely on the ability to learn the mappings between what actions to produce and when. This class of machine learning models found tremendous success in not only modeling animal's behavior but also in mapping its components and quantities to putative neural substrates, especially within BG architecture (Doya, 1999; K. N. Gurney et al., 2015; Maia, 2009).

1.5.3 Reinforcement learning

Following Thorndike's "Law of Effect" (Thorndike, 1911), reinforcement (RL) learning algorithms embody the intuition that an action followed by a positive outcome should be performed more often, *i.e.* reinforced. Such ability to learn when and what actions lead to good, or bad, outcomes presents an obvious evolutionary advantage. It allows agents to modify their expectations and, ultimately their behavior, in order to maximize the amount of resources that they are able to collect from the environment and hence their chances of survival. Interestingly, the same problems that animals face in the wild, and in the lab, are often also present when designing artificial intelligence (AI) agents. At the end of the day, an agent, be it biological or artificial, is inserted in an environment where it receives information from, and acts on, that environment. It has to make decisions on which actions to perform, under time pressure and uncertainty, and it should learn which actions to produce without the help of an ever-present knowledgeable teacher, but instead relying on iterative exploration of the environment. It is thus perhaps not surprising that AI algorithms designed to solve this problem also show great promise for modeling animal behavior. RL algorithms, in particular, have yielded great success in not only guiding AI research but also in providing a strong quantitative framework for animal behavior research. RL algorithms live somewhere between unsupervised (UL) and supervised learning (SL). While lacking the need for a labeled set of data to learn from, like in SL, it nevertheless has access to a teaching signal in the form of a *reward prediction error* (RPE). Specifically, RL relies on successive interaction episodes with the environment wherein an agent learns how to map states to actions (*policy*) with the end-goal of maximizing future returns. The nature of this interactive learning process, paired with a general reward function to be maximized, affords this class of algorithms its general usefulness.

As it will become apparent, quantities and modules present in some RL algorithms have clear neural isomorphisms, especially throughout basal ganglia. As a result, they represent, in my opinion, one of the most promising and exciting conceptual and quantitative frameworks to guide the research of animal behavior and its neuronal underpinnings.

1.5.3.1 Model-free and model-based controllers

Reinforcement learning algorithms can be further split into two large classes: model-free (MF) and model-based (MB). Agents employing model-based RL controllers have access to a model of the world, formally in the form of a state-transition matrix and a reward function. In other words, the agent has an internal representation of how states of the world relate to each other and how much reward to expect in a given state. Using this *"world model"*, the value of performing an action can be *bootstrapped* by planning (*i.e.,* simulating) using this "forward-model", and the policy producing the best outcome can be chosen. An obvious benefit of using a model of the environment is the ability to quickly adapt if the environment changes. For example, if before going home I am suddenly informed that my usual way back home is blocked, I can quickly plan, given a previously stored spatial map of the city, which alternative route to take. However, what makes these algorithms so powerful is often also their peril. Computationally, planning can be costly since for very deep or wide decision trees bootstrapping can take prohibitively long amounts of time when compared to the time-scale of a decision making process. Additionally, access to a complete model of the environment, especially for biological agents, is not always a very good assumption and thus, despite picking the optimal policy that will maximize returns given the agent's model of the environment, an incorrect, or otherwise incomplete model might result in suboptimal or maladaptive behavior.

MF agents, on the other hand, are not endowed with a model of the environment. They instead approximate the value of performing actions by directly interacting with the world, sampling its outcomes, and updating their expectations. Instead of computing actions-values through a forward-model as in MB, on each episode MF agents update the probability of performing a given action by comparing its present and expected outcome. If such outcome is better than expected, the action should be selected more often in the future, whereas actions that lead to worse than expected outcomes should have their probability of being taken decreased. Since these 'episodes' cannot be simulated and instead rely on directly acting in the environment, sampling strategies tend to be very

time-consuming, especially in situations with large action and state spaces that must be thoroughly explored in order to learn the optimal policy.

MF algorithms have been extensively applied to model animal's behavior and, depending on the implementation details of the state and action spaces, they turn out to exhibit surprisingly large expressiveness. Unfortunately, in many behavior tasks, it is somewhat difficult to ascertain whether animals are using a model of the world or simply a complex, perhaps very abstract, state-space representation. Recent research in AI have shown that neuronal networks trained with MF algorithms can exhibit MB-like behavior (Wang et al., 2018) and distinct deep-neural networks, also end-to-end trained with MF methods, are able to produce complex behaviors, like playing ATARI games, that we, as humans, often associate with need of an explicit model of the environment (Mnih et al., 2015). There are however examples where one would be hard pressed to explain animal behavior using MF controllers (Dayan & Daw, 2008; Doll et al., 2012; Daw et al., 2005a). As a result, the answer to which algorithms animals are employing is likely: "both". Under some scenarios, where behavior appears to be goal-directed, agents might leverage models of the world to guide prospective deliberation and simulate possible future scenarios. Similarly, in situations wherein animals have had little interaction with the environment, previously generated models, with some overlap with the current context, might be useful in that they might afford some degree of generalization. Alternatively, in situations where models are not useful or behavior becomes a simple contingency between stimulus and response (*i.e.* habitual), animals might instead rely on picking the action that has resulted in the largest return of rewards in the past, without any explicit knowledge of how that action will ultimately lead to reward.

Recent experimental data from selective activation or lesions in rodent striatum, suggest that these two controllers are implemented in distinct striatal subregions. Dorsolateral striatal circuits, enriched in sensorimotor information, have been found to be critical to establish model-free and habitual behavior whereas associative, dorsomedial striatum has been implicated in MB behavior. Yet, it is important to emphasize that correlates of MB computations and quantities have been found throughout many areas of the brain such as frontal cortical areas and

hippocampus. As a result, whether DMS is indeed implementing MB algorithms or simply has access to MB representations is an open debate.

Finally, it should be noted that these two modules likely operate in parallel, either by competing for behavior control or concertedly alternating given the specific demands of the task at hand. Indeed, several experiments have shown that for simple tasks MB and MF controllers are recruited serially. While initially, the goal-directed system seems to be engaged, as learning progresses and responses become more automatic, it is replaced by the habitual system. These results have also been corroborated by experimental lesions of DMS and DLS, respectively (Graybiel, 2008; Thorn et al., 2010; Gremel & Costa, 2013).

1.5.3.2 Temporal difference (TD) learning

In its simplest form, temporal difference learning (TD-learning) can be described as the trial-and-error process that agents use to learn the value of states of the environment leading up to reward. More precisely, the value of each state s_t, $V(s_t)$, will be given by the expected, discounted, sum of future rewards,

$$V(s_t) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t r_t | s_t = s], \tag{1.1}$$

Where γ is the discounted factor accounting for how much an agent devalues delayed reward in future states.

But how do agents learn these values through trial and error? At transitions where no reward is delivered, at each time step (*i.e.* transitioning from $t = 1$ to $t = 2$) the agent compares the value of its current state and the discounted value (by γ) of the next state ($\gamma \cdot V(s_{t+1}) - V(s_t)$). If the two perfectly cancel-out, the agent has learned the value of the stable environment. On the other hand, a mismatch between the two terms, or if a reward (r_t) is delivered, will indicate that there is still value to-be learned, and the animal updates $V(s_t)$ using the difference between what was expected and what it experienced. This difference

is called "reward-prediction error" (RPE), formally:

$$\delta_t = r_t + \gamma \cdot V(s_{t+1}) - V(s_t) \tag{1.2}$$

Where r_t the reward amount received at time t, $\gamma \cdot V(s_{t+1})$ the discounted value of the next state, and $V(s_t)$ the value of the current state. The value of state s_t, $V(s_t)$ is thus updated by,

$$V(s_t) \leftarrow V(s_t) + \alpha \cdot \delta_t, \tag{1.3}$$

Where α is the learning rate determining by how much $V(s_t)$ is updated. After several iterations, if the value function is perfectly learned, the RPE back-propagates to the earliest reward-predicting state thus signaling the arrival of a better than expected outcome. Likewise, in the absence of a predicted reward, a negative RPE is produced to signal a worse than expected outcome. As mentioned above, midbrain dopaminergic cells are thought to convey a quantity identical to a temporal-difference RPE in RL (Montague et al., 1996). Seminal electrophysiological experiments by Shultz and colleagues (Schultz et al., 1993, 1997) have shown that, similarly to the RPE in RL models, primate DA neurons fire to unpredicted rewards. Moreover, after several pairings of a cue to the delivery of the reinforcer, the response that was initially locked to unexpected reward delivery, now shifts to the earliest reward-predicting cue. Finally, in striking accordance with RL theory, omitting the reward after cue delivery results in brief pauses of DAergic neural activity around the time of expected reward delivery. After these initial observations, other theoretically predicted features of RPE have been observed in the activity of midbrain dopamine neurons. Among others, DAergic neurons were shown to be sensitive to reward magnitude and probability (Fiorillo et al., 2003), as well as to the delay between the conditioned (CS) and unconditioned stimulus (US) (Roesch et al., 2007). However, more recent experimental data has been used to argue that dopaminergic neurons might code more than the *classic* RPE. For one, what was once thought to be a scalar, one-dimensional, RPE signal broadcasted to the whole striatum, it has now been

shown to be different across nearby dopaminergic cells with respect to relative sensitivity to positive and negative RPEs (Dabney et al., 2020), and to qualitatively vary depending on the midbrain source and output striatal areas. For example, DAergic axon terminals in the tail of the striatum have been shown to respond phasically to threatening stimuli - which intuitively should produce a negative RPE (Menegas et al., 2018). Moreover, optogenetic activation of these terminals was shown to produce avoidance rather than appetitive conditioning.

In another set of examples, dopamine cells in SNc, that project to dorsal regions of the striatum, have been shown to respond to not only reward-predicting cues but also to the onset of actions (Howe & Dombeck, 2016; da Silva et al., 2018; Coddington & Dudman, 2018), and cells in VTA have been shown to respond not only to rewards but also to features of the environment and animals' own behavior (*e.g.,* velocity, acceleration) (Engelhard et al., 2019). While it is clear that these results do not immediately fit in the originally proposed RL theory of BG function, they can be accommodated by, once again, considering the parallel functional architecture of the BG circuitry. It is thus possible that these signals do not reflect a different computation (*i.e.,* RPE) but instead the same computation performed on top of distinct inputs with access to different information (Lau et al., 2017). Specifically, in the case of the inputs arriving at dorsal striatum from SNc, it is definitely possible that responses aligned to movement reflect the reward expectation learned from performing those same movements (Gadagkar et al., 2016), which can be used to construct a state-space representation on top of which values can be learned (R. Chen & Goldberg, 2020).

TD-learning algorithms continue to yield great success in explaining and modeling simple learning processes like classical conditioning, where agents do not explicitly need to produce actions. Behavior, however, extensively relies on not only making predictions, but also acting on these. In order to accommodate this requisite, we need to extend this framework such that it can also learn *what* and *when* actions should be produced.

1.5.3.3 Actor-critic architecture

In order to learn what *to do* in each state (*policy*) we modify the previous algorithm to include actions. A simple extension can be made by simply replacing the value of a state s_t, $V(s_t)$ by a value of a state-action pair, or Q-value, $Q(s_t, a_t)$. Similarly to TD-learning, the Q-value function is then learned through successive episodes wherein the agent is able to produce actions, following a given policy π, and experience outcomes. After exploring the environment and having learned the value of the various state-actions pairs, the agent will be able to choose an action that maximizes $Q^{\pi}(s_t, a)$for a given state s_t. This algorithm is referred to as Q-Learning (Sutton et al., 1998).

This framework presents a few challenges. For one, Q-learning cannot explicitly learn stochastic policies (*e.g.,* sometimes the optimal strategy is to not always perform the same action in a given state). Second, some RL problems, especially in the continuous domain, become intractable when learning $Q(s_t, a_t)$ pairs directly, since the number of actions, for any given state, might be too large. For these types of problems, policy-based learning has been shown to be more adequate (Sutton et al., 1998). Actor-critic (AC) architecture marries the two approaches at the cost of some added complexity. As the name implies, AC models rely on two distinct functional modules: the *Actor* and the *Critic*. In its simplest form, the goal of the *Actor* is to learn the best policy,

$$\pi_t(a|s) = P\{a_t = a|s_t = s\} \tag{1.4}$$

that leads to the largest returns. The feedback that the *Actor* uses to update its policy is provided by the *Critic* in the form of an RPE which, similar to a TD-learning agent, aims to maximize the weighted sum of future rewards. The key difference being that this adaptive *Critic* module learns a value function ($V^{\pi}(s_t)$) that is not only dependent on its state representation but also on the policy the *Actor* is currently following.

In addition to the aforementioned neural isomorphism found between dopaminergic neuron phasic activity and RPE, AC models find an additional parallelism between cortico-striatal synapses and the *policy*. Under this frame-

work, synapses between cortical inputs, carrying information about the current state of the environment and available actions, and medium spiny neurons store the policy. In AC architecture, action preferences are continuously updated using the error provided by the adaptive critic. Similarly, cortico-striatal synapses have been shown to be modulated by the RPE-carrying signal provided by striatal dopaminergic input, which has been shown across different experimental paradigms to be capable of reinforcing state-action contingencies (Frank et al., 2004; Bromberg-Martin et al., 2010; Tai et al., 2012; Danjo et al., 2014; Yttri & Dudman, 2016). Plasticity dynamics at cortico-striatal synapses significantly differ between the two BG pathways (Shen et al., 2008; Yagishita et al., 2014; Iino et al., 2020; S. J. Lee et al., 2021). While direct pathway medium spiny neurons undergo LTP with increases in extracellular dopamine concentrations and LTD, with drops below baseline levels, indirect pathway MSNs respond in the diametrically opposed manner. This sets up a situation wherein dMSNs appear to *learn* from positive reward prediction errors (*i.e.* when things turn out better than anticipated) while iMSNs learn from situations that result in worse than expected outcomes. As I will suggest in chapter 2, this could be a neuronal substrate responsible for associating cortical state/action information to *policy* by essentially mapping the tendency to perform an action on a given state, $P\{a_t = a|s_t = s\}$, to the synaptic strength between BG input and medium spiny neurons.

Finally, the *actor* and *critic* modules have also been mapped onto different structures within BG. In one set of models (Bornstein & Daw, 2011), dorsal striatum (DS), where correlates of action value (Hikosaka et al., 2006; Lau & Glimcher, 2007, 2008; H. F. Kim & Hikosaka, 2013) and instrumental prediction errors have been found, was suggested to implement the *actor* module. The role of the *critic*, on the other hand, was assigned to ventral striatum where activity consistent with pavlovian reward prediction errors was reported (Cardinal et al., 2002; O'Doherty et al., 2004). This model is supported by interesting observations wherein some violations to the general parallel architecture of BG have been found. For instance, while nigrostriatal connections are to a large extent

reciprocal, ventral areas of SNc - that are known to receive input from ventral striatal regions - are known to harborize to DLS (Haber, 2014).

A second class of models have tried to map this function to other known anatomical features of the BG circuit. Under these models, actor-critic architecture within DS is implemented in functional and anatomically segregated compartments named *striosomes (or patches)* and *striatal matrix* that were proposed to implement the adaptive *critic* and the *actor* modules, respectively (Houk et al., 1994). The appeal of this theory rests on the anatomical and physiological findings that MSNs in the patch compartments project to DAergic midbrain areas, whereas those in the matrix project to SNr/GPi (Gerfen, 1984; Fujiyama et al., 2011; Watabe-Uchida et al., 2012), providing the information necessary for the computation of a critic-derived RPE and action values, respectively. Some authors have nonetheless pointed out that some of the anatomical features found in primate and rodent BG are not consistent with this model (Joel et al., 2002).

At the end of the day, many questions about the exact implementation details of AC architecture in BG remain unanswered. It is likely that BG and AC will never find a perfect mapping in structure and function. Nevertheless, the fact that it guided our conceptual understanding of the circuitry cannot be understated and thus remains, in my opinion, one of the most complete and successful models of BG function.

1.6 Outro

In light of this experimental and theoretical backdrop, several questions pertaining to the role of BG in behavior remain to be answered. The next chapter will present experimental data collected from experiments designed to fill some of these knowledge gaps.

While the two pathways have been shown to bidirectionally regulate the output function of BG, and influence behavior, co-activation is often observed during movements. How can one reconcile these findings? I will argue that, under a conceptual model where both suppression and promotion aspects of action selection

are necessary to generate adaptive behavior, co-activation is indeed expected to be observed.

Additionally, I will describe a newly developed behavior paradigm during which sensorimotor indirect pathway is shown to be actively engaged during periods wherein animals must suppress a potentially rewarded, and thus perhaps tempting, action. A feature of the indirect pathway function often left experimentally unprobed.

Finally, chapter 2 will conclude with a computational model to explain how behavior, neuronal dynamics, and manipulations can arise by considering a multi-agent RL architecture. In this model, two parallel systems with distinct *views* on the world interact, through behavior, to generate a general behavior policy.

These findings will be discussed, and expanded on, in the general discussion chapter of this thesis.

Chapter 2

Regionally distinct striatal circuits support broadly opponent aspects of action suppression and production

2.1 Introduction

Adaptive behavior involves a judicious combination of suppression and production of actions. A predator must suppress its urge to pounce until its prey is within reach, just as humans must suppress giving in to temptation to secure longer-term rewards.

The basal ganglia (BG) are a collection of subcortical structures (Albin et al., 1989) that are thought to regulate the appropriate selection of actions depending on expected consequences(Schultz, 1995; Doya, 1999). In addition, the inability to balance action production and suppression is associated with disorders that involve the BG such as ADHD(Barkley, 1997), Parkinson's, and Huntington's diseases(Albin et al., 1989). Interestingly, two major BG circuits, the so-called direct and indirect pathways, possess anatomical and molecular characteristics consistent with promoting and suppressing actions, respectively (Gerfen & Surmeier, 2011; Alexander & Crutcher, 1990). These two pathways originate in the major input area of the BG, the striatum, at direct striatonigral medium spiny neurons (dMSNs) and indirect striatopallidal medium spiny neurons (iMSNs) that project directly or indirectly toward the output areas of the BG. While multiple lines of evidence suggest functional opponency between the two pathways, an apparent discordance between neural activity on the one hand, and anatomy and cell type-specific perturbation data on the other has led to ongoing debate regarding the rules that govern BG circuit function.

As predicted by anatomy (Y. Smith et al., 1998), activating dMSNs can rapidly suppress, while activating iMSNs can rapidly enhance, the activity of inhibitory output neurons of the BG in the substantia nigra (Deniau & Chevalier, 1985; Kravitz et al., 2010; Freeze et al., 2013). At a behavioral level, activation of dMSNs consistently produces opposite effects to those of activating iMSNs with respect to locomotion(Kravitz et al., 2010; Roseberry et al., 2016), ongoing motor sequence production(Sippy et al., 2015; Tecuapetla et al., 2016), reinforcement (Kravitz et al., 2012; Yttri & Dudman, 2016), and value-based decisions (Tai et al., 2012). However, the activity of dMSNs and iMSNs in sensorimotor striatum appears to be largely positively correlated around action production (Cui et al.,

2013; Tecuapetla et al., 2014; Barbera et al., 2016; Markowitz et al., 2018). Such observations of concurrent activation of the two pathways have been used to argue against the hypothesis that they functionally oppose each other (Cui et al., 2013; Tecuapetla et al., 2014; Cox & Witten, 2019).

A longstanding, and potentially reconciling, view of BG circuit function is that the two pathways might contribute to selection amongst various actions in a competitive manner (Denny-Brown & Yanagisawa, 1976; Mink, 1996; Redgrave et al., 1999). In this view, action selection proceeds through combined promotion of motor programs by the direct pathway, and suppression of motor programs by the indirect pathway. Such a model predicts broad coactivation of the two pathways during action production even as they function in opposition to each other. Notably, this framework also predicts that sustained suppression of action should promote large-scale decorrelation or even anticorrelation between the two pathways. This possibility, to our knowledge, remains untested. Activity of iMSNs should be elevated to suppress action, while the activity of dMSNs should be limited until action is released. Such observations would naturally reconcile currently disparate interpretations of BG circuit function.

To test this hypothesis, we employed a variant of an interval categorization task (Gouvêa et al., 2015; Soares et al., 2016) that requires a series of self-initiated and cued actions, and critically, a sustained period of dynamic action suppression. During this behavior, we then recorded activity from dMSNs and iMSNs in the dorsolateral striatum (DLS) of mice using fiber photometry and electrophysiology. We found that both pathways displayed phasic activation during action production, as previously reported. However, action suppression revealed clear differences in activity and signatures of functional opponency, most prominently in the form of privileged engagement of iMSNs and with opposite dynamics as compared to dMSNs.

To assess the functional importance of the observed patterns of neural activity, we then performed a series of optogenetic inhibition experiments. The results provided further support for the idea that the direct and indirect pathway support broadly opponent function but, surprisingly, revealed little engagement of DLS

circuits in the promotion of actions. Instead, DLS appeared to be largely engaged to suppress a given action when it was most tempting.

A dominant view holds that BG circuits function similarly to components of reinforcement learning (RL) models, learning through experience to apply a program for behavioral control that maximises reward (Doya, 1999). In this view, diverse inputs from a broad range of cortical and thalamic areas provide striatal circuits with information about the state of the world, and dense dopaminergic inputs carrying reward prediction errors cause synaptic plasticity and thus teach the system a mapping of which actions to take in any given state. This mapping constitutes a policy for behavioral control that exerts its influence through projections from BG output nuclei to motor circuits. But a local policy of selectively suppressing actions that are tempting, such as that we found in DLS, requires a circuit elsewhere to provide the temptation itself.

It has long been appreciated that the BG possess a parallel circuit architecture. Sensorimotor, associative, cognitive and limbic-related information passes through BG circuits in a largely segregated manner (Alexander & Crutcher, 1990). Based on this functional anatomy, it stands to reason that any policy-related information learned in the striatum will differ depending on which parallel circuits are involved. Thus, we hypothesized that the suppressive function observed in DLS may be learned as a consequence of an action promoting policy learned by parallel circuits involving other regions of the striatum. We then constructed a simplified reinforcement learning model that shows how DLS circuits can interact through behavior with parallel circuits located elsewhere to produce the observed patterns of data. The model reproduced behavior, pathway specific patterns of DLS neural activity, and differential effects of inhibiting iMSNs on the two hemispheres. It also made an implicit prediction: that some other region of the striatum promotes the actions that DLS has learned to suppress. To test this prediction, we performed an additional set of optogenetic experiments in more associative, dorso-medial striatum (DMS). DMS has been implicated in more forward-looking, "model-based" behavioral control (Bornstein & Daw, 2011), and we reasoned that the basis for temptation in the current task might lie in the consideration of the future states wherein particular actions would result in reward.

Inhibiting either dMSNs or iMSNs in DMS had no detectable effect on action suppression, however, unlike in DLS, unilaterally inhibiting dMSNs in DMS did reduce the production of contralateral actions, consistent with the predicted role for higher order circuits in promoting action.

More generally, the model illustrates how specific BG circuits may give rise to distinct influences on behavioral control that depend not only on the environment, but on the impact of other circuits that learn and exert their influence in parallel. This highlights an underappreciated mode of behavioral control whereby the brain functions much like a multi-agent system, with a variety of concurrent influences on behavior, sometimes aligned (Cartoni et al., 2016)and sometimes conflicting (Dayan et al., 2006). Together, these findings provide new insight into how BG circuitry can contribute to distinct aspects of action production and suppression across different striatal territories, with broad implications for understanding the neural mechanisms of both normal and pathological behavioral control.

2.2 Results

2.2.1 Production and proactive suppression of action

We trained mice on a variant of a two-alternative time interval categorization task wherein subjects were required to suppress movements during interval presentation. Briefly, mice self-initiated a trial by inserting their snout into a centrally located initiation nose port, eliciting a brief auditory tone (Figure 2.1). Mice were then required to maintain their position in the initiation port (*fixation*) until a second auditory tone was delivered. This second tone was delivered at a delay that was randomly chosen from a set of 6 intervals, symmetric about 1.5s, and ranging from 0.6s to 2.4s. After delivery of the second tone, animals were free to choose either of two choice ports located at an equal distance to either side of the initiation port. Rewards were delivered for choices to one side ("short" choice) if the presented interval was shorter than a 1.5s decision boundary, and at the opposite choice port ("long" choice) if the interval was longer than 1.5s. Mice learned to categorize interval stimuli much longer or shorter than the decision-boundary with high accuracy (92.1±0.7%, 0.6s and 2.4s intervals, mean±s.e.m.

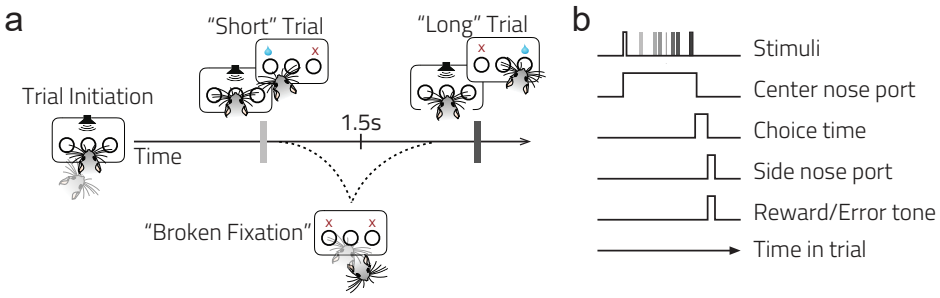


Figure 2.1. Task **(a)** and event **(b)** diagram. Subjects self-initiate each trial in a center nose port. After a variable delay a second tone is played and they are asked to categorize the presented interval as "short" or "long" by responding in one of two side ports. Between the two tones subjects are required to maintain position in the centre port - "fixation".

n = 14 mice), yet choices were more variable for intervals nearer to the decision-boundary, Figure 2.2a). In addition, mice produced a stereotyped movement profile over each trial (Figure 2.2b). Movement speed increased leading to trial initiation, followed by a brief period of postural adjustment before animals settled into a period of immobility until the second tone was delivered. Immediately following second tone delivery, movement speed increased again as animals executed their choices. If animals failed to maintain fixation in the initiation port until the second tone, an error tone was immediately delivered and the trial was terminated (36.5±2.1% of all trials, n=14 mice). We will refer to these trials as *broken fixations* throughout the text. Interestingly, animals often entered a choice port even after *breaking fixation* and aborting the trial (52.1±4.9% of *broken fixation* trials, n=14 mice). These choices were executed with a similar timecourse and trajectory as valid choices (Figure 2.3) and were largely "appropriate", toward the "short" choice port when breaking early in a trial, and toward the "long" port when breaking late in a trial (Figure 2.4a-b). The pattern of *broken fixations* reflects the overall reward associated with the two choices over time, and not the likelihood of second tone occurrence (Figure 2.2a, Figure 2.4c). These data are consistent with animals developing a dynamic motor plan that remains latent as long as it is successfully suppressed. Failure to suppress this temptation led to premature execution of the planned action.

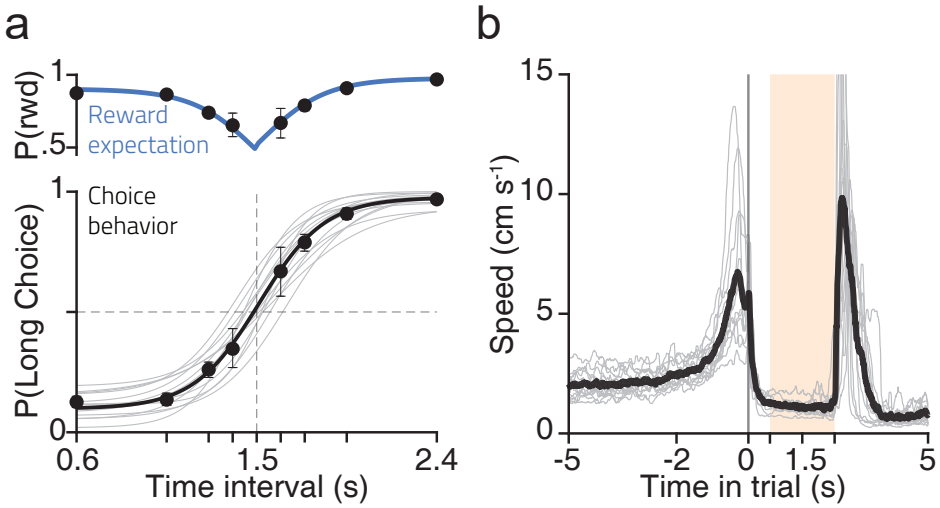


Figure 2.2. **a)** Bottom: Psychometric fit to the performance of each mouse that underwent photometric recordings (n=14, light gray) and the logistic fit to the overall average performance across mice (black). Top - Reward expectancy calculated from the overall performance of animals on a given stimuli (blue trace depicts the rectified psychometric fit at 1.5s) **b)** Animal's head speed as a function of time, aligned on trial initiation, for a single random session of each mouse that underwent photometric recordings, during trials wherein the longest interval (2.4 seconds) was delivered and a correct choice performed. Gray, mean of individual animals; black, average of all mice (n=14). Shaded region highlights period of immobility (0.6s to 2.4s post-trial initiation). Error bars represent s.e.m..

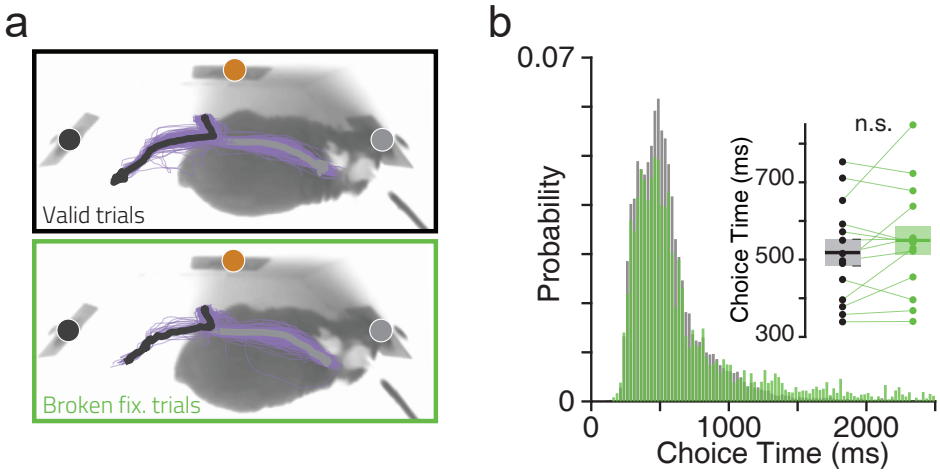


Figure 2.3. **a)** Comparison of average nape trajectories, during a choice movement (-0.5 to 1.5 seconds relative to leaving the center port) , for sessions of a single animal, for trials wherein animals chose the short (black) or long (gray) nose port on valid (top, black outline) or broken fixation (bottom, green outline) trials. Thinner purple lines depict single trials and circles represent the Long (Black circle), Initiation (Orange circle) and Short (Gray circle) nose ports (see also Figure 2.24); **b)** Distribution of choice times, i.e. time taken for the animal to leave the centre port and report its choice, in completed trials (black) and broken fixation trials (green). Inset depicts the medians of choice times per animal for completed and broken fixation trials (one-sample t-test, P = 0.155, t_{13} = 1.511). Error bars represent s.e.m..

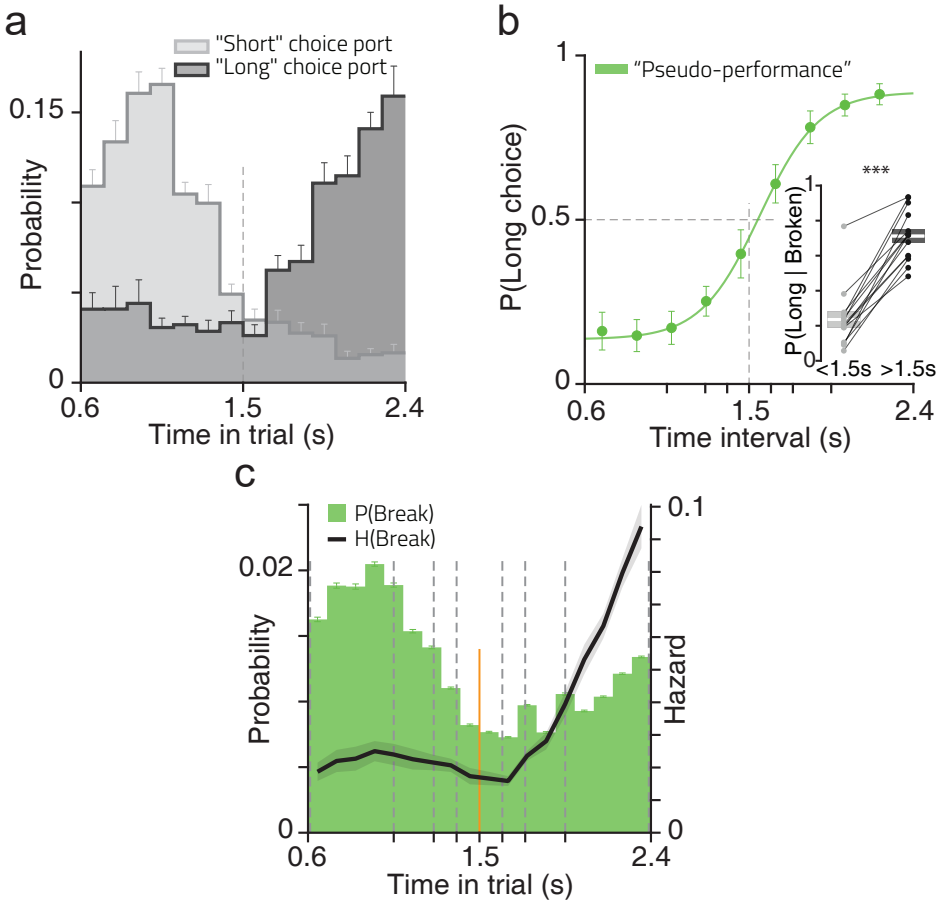


Figure 2.4. **a)** Probability density functions of broken fixations over time during the immobility period (0.6s to 2.4s), contingent on subsequent choice at one of the side ports. **b)** Average overall "pseudo-performance" of all animals used in the photometric recordings calculated from broken fixation trials. To calculate the performance in broken fixation trials, we binned the times at which animals aborted the trial and calculated the proportion of reports at the "long choice" port over all reports. Inset depicts single animal probability of choosing long as a function of breaking fixation before ($<$1.5s) or after ($>$1.5s) the decision boundary (one-sample t-test, P$\ll$0.001, $t_{13} = 11.178$). **c)** Probability density function of Broken fixation occurrence, over all trials, as a function of time since first tone (green) and corresponding hazard rate (black full line). Grey dashed lines represent times at which a second tone might occur. Orange full line represents the decision boundary (1.5s). Error bars represent s.e.m..

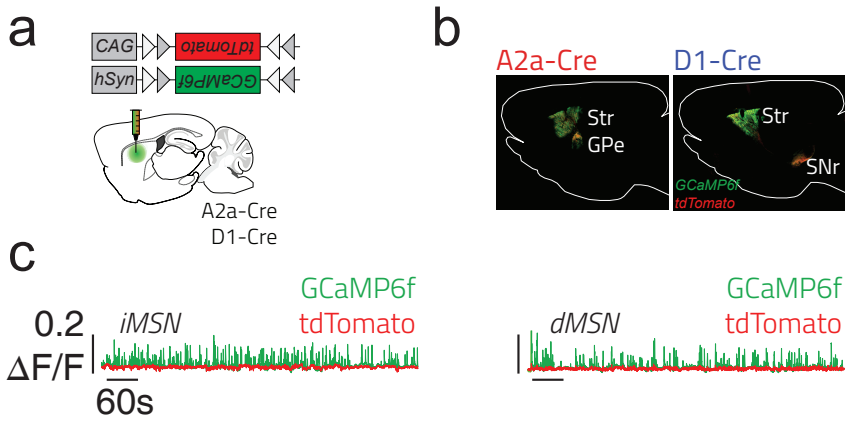


Figure 2.5. **a)** Viral strategy used to record Ca2+ activity in dorsal striatum. **b)** Pattern of transgene expression in A2a-Cre (iMSN) or D1-Cre (dMSN) animals in sagittal section, 2.1mm ML. Str-Striatum, GPe-Lateral globus pallidus, SNr-Substantia nigra pars reticulata. **c)** example of photometric traces in indirect (left) and direct (right) pathway MSNs.

2.2.2 Opposite modulation of striatal direct and indirect pathways during action suppression

To probe the large-scale activity of the direct and indirect pathways for signs of functional opponency during movement and active suppression of lateralized movements, we recorded the activity of dMSNs and iMSNs in the dorsolateral striatum during task performance. We combined mouse lines expressing Cre recombinase in either dMSNs (D1-Cre EY217Gsat line) or iMSNs (A2a-Cre, KG139Gsat line) with cre-dependent viral expression of the calcium indicator GCaMP6f(T.-W. Chen et al., 2013)(Figure 2.5), using coordinates where coactive dMSNs and iMSNs have been described during movement (Tecuapetla et al., 2014)(Figure 2.6). No significant differences in behavior were detected between the two mouse lines (Figure 2.7, Figure 2.8). We then used fiber photometry ((Soares et al., 2016; Matias et al., 2017), Figure 2.5, Figure 2.9) to access the pooled activity of a local population of dMSNs or iMSNs in the dorsolateral striatum. To determine whether gross differences in activity patterns were present across a trial, we first examined the combined activity of neurons located in both

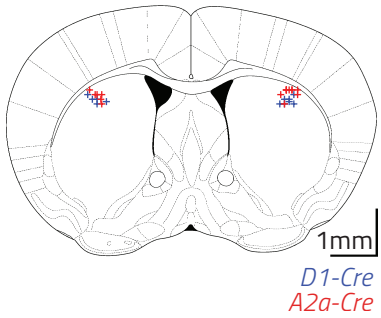


Figure 2.6. Histological reconstruction of sites of fiber implantation for photometry experiments in DLS. Animals are colored by their genotype according to the legend. All coordinates were projected to the same coronal slice (AP = +0.5 from bregma) adapted from Franklin & Paxinos (2008).

hemispheres. In individual animals, we observed that activity of either dMSNs or iMSNs increased around task epochs when animals were required to take action, namely trial initiation and choice execution, consistent with previously observed coactivation of the two pathways (Figure 2.9). Indeed, across all animals, though the time courses of activity appeared to differ slightly, the mean activity of the two pathways was indistinguishable around trial initiation Figure 2.9, iMSN vs dMSN = 0.141 [-1.09, 1.37] zΔF/F, p = 0.8065; Effect Size, 95%[CI], p-value). However, activity in dMSNs and iMSNs displayed marked differences during interval presentation, when mice were required to suppress movement (Figure 2.9). Across all animals, iMSN activity was significantly elevated relative to dMSN activity (Figure 2.9, iMSN>dMSN = 0.783 [-0.071, 1.64] zΔF/F, p = 0.0345). We confirmed these overall patterns of activity by recording single units electrophysiologically from DLS and optogenetically photo-identifying iMSNs(Lima et al., 2009). We observed diverse patterns of responses in both photo-identified, and non-photo-identified putative MSNs (Figure 2.10, Figure 2.11) (Jin & Costa, 2010; Klaus et al., 2017). Consistent with photometry, the photo-identified iMSN population was significantly enriched for cells displaying increased activity during action suppression (Figure 2.11). The difference in photometric signals recorded from iMSNs and dMSNs was not constant, but grew on average as mice were required to suppress action for longer periods. We wondered whether the need

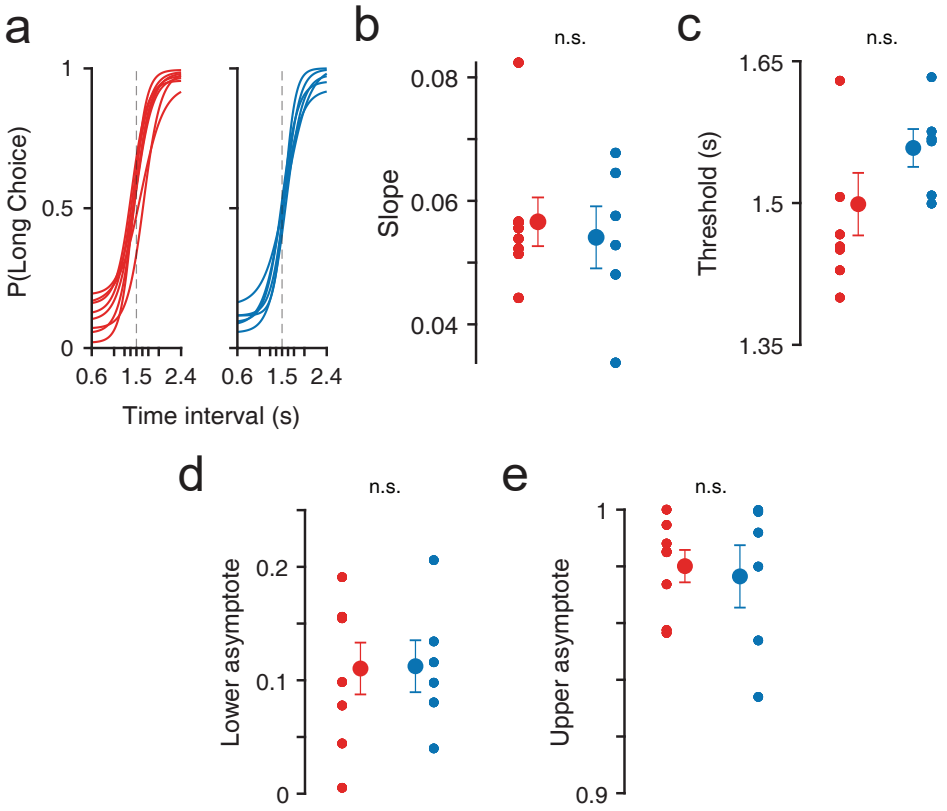


Figure 2.7. No significant differences in psychometric behavior were detected between genotypes. A2a-Cre and D1-Cre single animals, included in the photometry experiments, are shown in red and blue, respectively. a-e) Single animal psychometric curve **(a)** fits and respective parameters (see Methods for further details, two-sample t-test, b) P = 0.935, t_{12} = 0.083, c) P = 0.17, t_{12} = -1.459, d) P = 0.823, t_{12} = 0.228, e) P = 0.826, t_{12} = -0.225). Error bars represent s.e.m.. n.s. P>0.05.

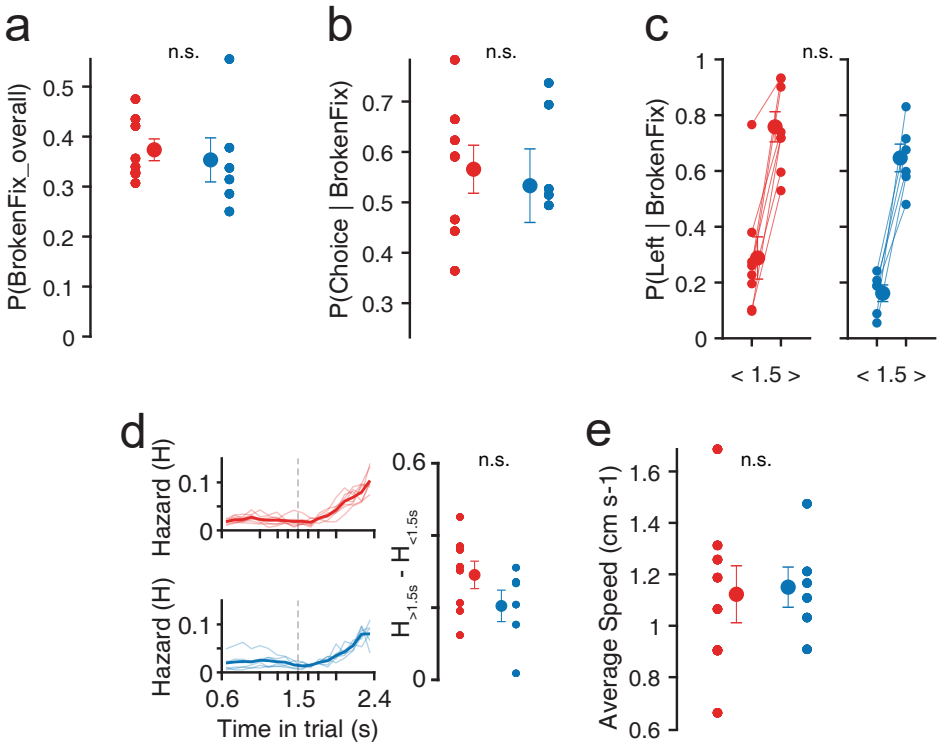


Figure 2.8. No significant differences in behavior, in broken fixation trials and movement during the delay period, were detected between genotypes. A2a-Cre and D1-Cre single animals, included in the photometry experiments, are shown in red and blue, respectively. **a)** Overall probability of breaking fixation (all trials included) (P = 0.665, t_{12} = 0.445) **b)** Percentage of trials wherein animals attempted to make a choice after breaking fixation (all trials included) (P = 0.703, t_{12} = 0.391). **c)** Probability of reporting at the "long choice" port after breaking fixation contingent on whether the animal aborted before (<-1.5s) or after (>1.5s) the decision boundary (P = 0.872, t_{12} = -0.165). **d)** Left, Hazard of breaking fixation in time for single animals (thin curves) and overall averages within genotype (thick lines). Right, differences between the hazard of breaking fixation after and before the decision boundary(P = 0.165, t12 = 1.48). **e)** Mean velocity during the delay period from correct trials of the longest interval (2.4 seconds, Data from Figure 2.2, P = 0.892, t_{12} = -0.139). Error bars represent s.e.m.. n.s. P>0.05.

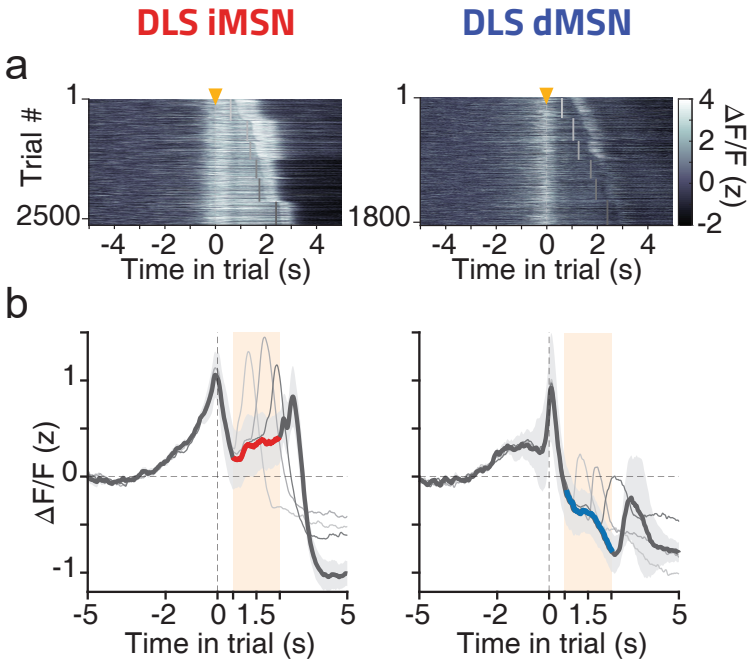


Figure 2.9. **a)** Single-trial photometric data (z-scored, see methods) for all correct trials of a single A2a-cre (left) and D1-cre (right) animal across all sessions aligned to trial initiation (yellow arrow). Interval offset is represented as vertical grey bars, where darker grey represents longer intervals. Trials were further ordered within interval by reaction time. **b)** Average activity (z-scored) across all animals of a given genotype (A2a-Cre n=8, D1-Cre n=6). Darkest trace represents activity during the longest interval within the interval set (2.4 seconds) and lighter gray traces corresponding to a subset of shorter intervals. Colored segments of the trace highlight a period of immobility (0.6s to 2.4s post trial initiation). Error bars / boxes represent s.e.m.

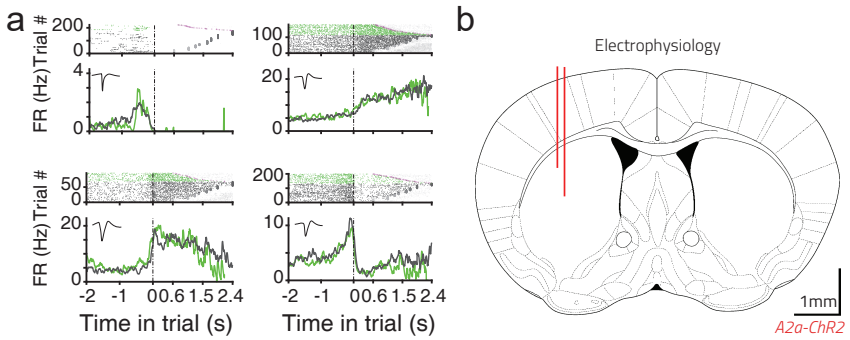


Figure 2.10. **a)**PSTHs of four representative photo-identified units, aligned to trial initiation, exhibiting distinct patterns of activity during the delay period. For each units, top shows a raster plot and bottom the corresponding PSTH. Black and green depict valid and broken fixation trials respectively (averaged during relevant epochs, i.e. before second tone or broken fixation, respectively). **b)** Histological reconstruction of sites of optrode implantation (tapered fiber) used for electrophysiological recordings and iMSN photoidentification. All coordinates were projected to the same coronal slice (AP = +0.5 from bregma) adapted from Franklin & Paxinos (2008).

to suppress action might similarly grow over time. We computed the probability that mice *break fixation* at each time bin within the delay period conditioned on their not having *broken fixation* up to that point, a quantity known as the hazard rate. In the context of this task, computing the hazard rate as opposed to the overall probability of *breaking fixation* controls for the fact that animals experienced more instances of early time bins in the delay and thus had more opportunities to *break fixation* early. After a brief dip in *broken fixation* around the decision boundary, the hazard rate of overall *broken fixation* behavior rises dramatically (Figure 2.4c). We next examined the hazard rates of *broken fixation* conditioned on subsequent choice, and found that the early mode in the hazard rate of overall *broken fixations* was comprised of trials where the mice subsequently made a short choice, whereas the late rise in the overall hazard was comprised of trials where mice subsequently made a long choice (Figure 2.12). Notably, the tendency to *break fixation* in a particular direction was asymmetric. The urge to *break fixation* and make long choices late in the trial appears to far

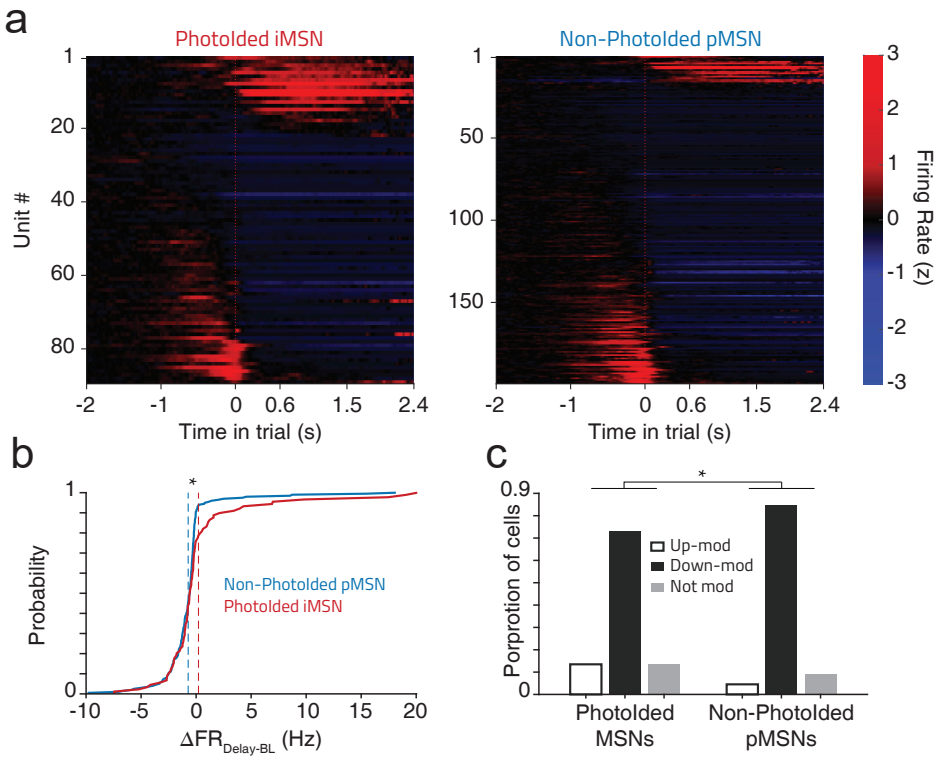


Figure 2.11. Photo-identified iMSNs population is enriched in cells with higher firing rate during the delay period. **a)** Activity profile of photo-identified indirect pathway medium spiny neurons (photo-Ided iMSNs) and non-photo-identified putative MSNs (see methods for details). Each row represents a unit's z-scored activity aligned to trial initiation that results from averaging the activity for all intervals cropped at second tone. Units are ordered by the angular position formed by the first two principal component projections. PCs were computed using a period of -2 to 2.4 seconds from trial initiation. **b)** Cumulative distribution of changes in firing rate during the delay period of photo-Ided iMSN (red) and all other putative MSNs (blue). Average ΔFR is significantly larger for iMSNs when compared to the distribution of non-identified cells (two-sample t-test, P = 0.0196, $t_{286} = 2.35$). **c)** Proportion of up, down and not modulated cells during the delay period (see methods for details). Proportions are significantly different between the two groups (Chi-squared test, P = 0.0115, $\chi^2_2 = 8.939$).

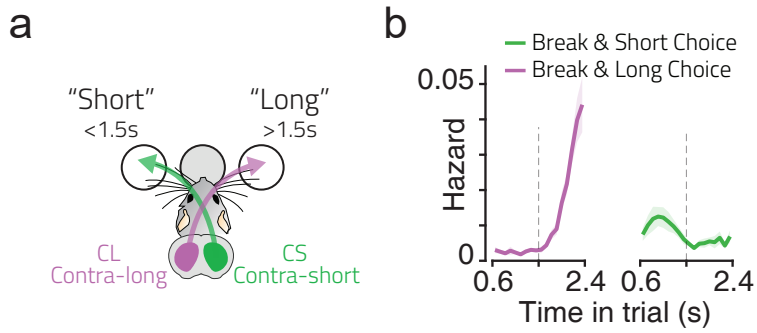


Figure 2.12. **a)** Schematic of labelling convention, the three large circles represent the three nose ports present in the behavioral task apparatus. The filled grey circle represents the trial initiation port, while the open circles represent the two choice ports. **b)** Hazard rate of broken fixations trials (see methods) wherein animals subsequently made a choice at the port corresponding to a short (green, right) or long (purple, left) choice. Error bars / boxes represent s.e.m.

outweigh the urge to *break fixation* and make short choices early in the delay period. This asymmetry may reflect the increasing certainty that animals gain over time about the ultimate location of reward at the "long" choice port. If MSN activity during successfully completed trials acted to suppress these lateralized urges to act early and late, we might expect differences in the time course of activity between the two hemispheres. Indeed, in the hemisphere contralateral to the rewarded location for "long" stimuli (contra-long, CL, Figure 2.12a) iMSN and dMSN activity steadily increased and decreased throughout the delay period, respectively (Figure 2.13a, difference between pre and post decision boundary mean activity: iMSN:CL = 0.423 [0.006 0.840]zΔF/F, p = 0.0462, dMSN:CL -0.535 [-1.017 -0.054]zΔF/F, p = 0.0272).

In contrast, in the hemisphere contralateral to the rewarded location for "short" stimuli (contra-short, CS, Figure 2.12a) activity levels in both pathways were relatively constant as compared to the CL hemisphere (Figure 2.13a, iMSN:CS = -0.104 [0.521 0.313]zΔF/F, p = 0.928, dMSN:CS -0.040[-0.052 0.442]zΔF/F, p = 0.999)). These data reflect a situation where lateralized patterns of activity reflected the strength of an urge to move contralaterally over

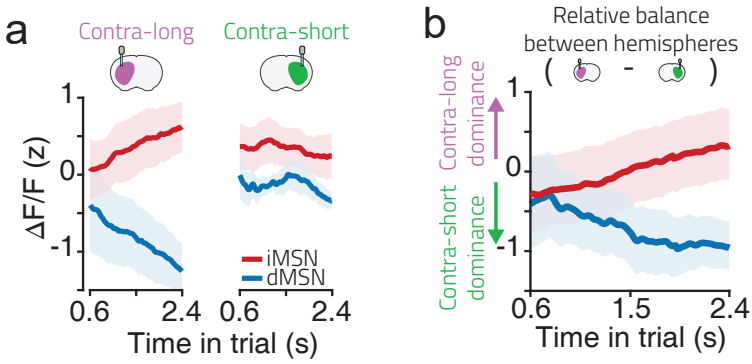


Figure 2.13. **a)** Averaged normalized activity recorded from the hemisphere contra-lateral to a long (left panel) or short choice (right panel) port (z-scored) across all iMSN (red) and dMSN (blue) mice. Only correct completed trials were included.**b)** Average of all pairwise differences in immobility period activity of dMSNs and iMSNs between the two hemispheres, subtracting activity recorded in hemispheres contralateral to the "long" choice port from activity recorded in hemispheres contralateral to the "short" choice port (i.e., CL activity - CS activity). Error bars / boxes represent s.e.m.

time. Relative levels of activity between the two hemispheres varied over time, and in opposite directions in the two pathways, in accordance with behavior (Figure 2.13b). Such observations may indicate that BG circuitry residing in a particular hemisphere is preferentially recruited to suppress movements to the contralateral direction when and to the degree that the animal is tempted to move in that direction.

The pattern of pathway specific activity observed in the two hemispheres over time suggests that DLS iMSNs are engaged to dynamically suppress the temptation to act. Consistent with this, we detected significant downward deviations in the rate of change of photometric signal(Markowitz et al., 2018) in iMSNs preceding *broken fixations* as compared to time-matched control periods (Figure 2.14a). Photo-identified iMSNs recorded electrophysiologically that were engaged during action suppression also displayed significant decreases in firing rate preceding the movement (Figure 2.14c).

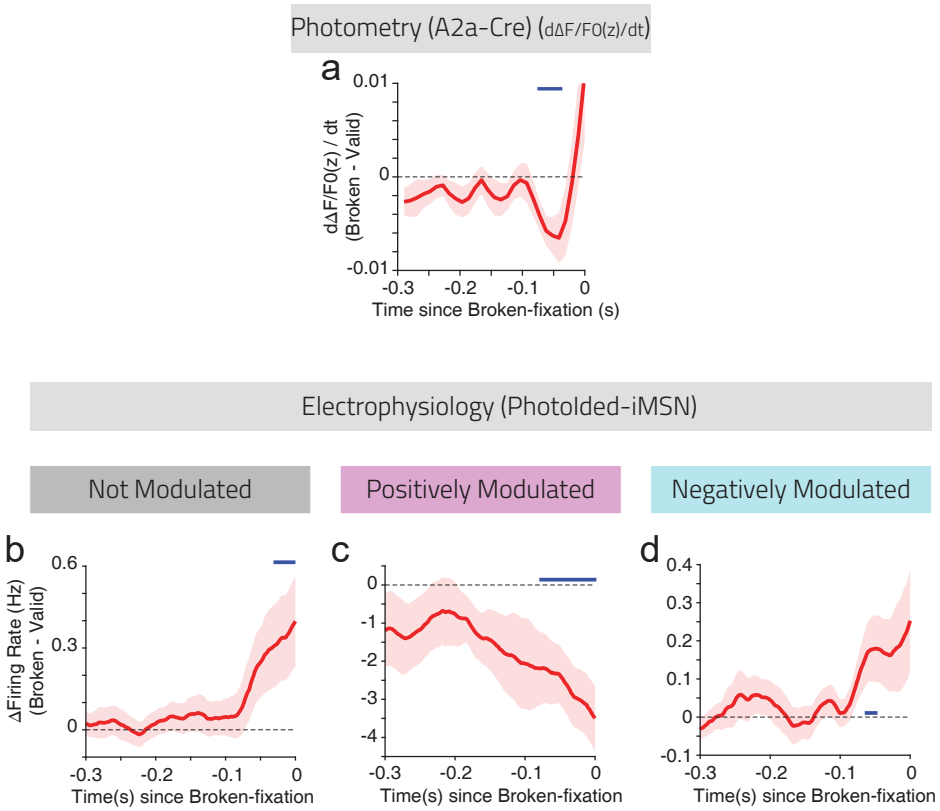


Figure 2.14. **a)** Difference of the rate of change (derivative, $d(\Delta F/F_{BrokenFixation})/dt - d(\Delta F/F_{Valid})/dt$) of photometry signal for hemispheres recorded in A2a-Cre animals (n = 16 Hemispheres), and difference of mean activity ($FR_{BrokenFixation} - FR_{Valid}$, Hz) of all non-modulated (**b**), positively modulated ()**c**) or negatively modulated (**d**) photo-identified iMSNs (n = 46, 12 and 31 units, respectively, from 2 animals), aligned to the time of broken fixations (for details on the analysis see methods, Full lines indicate the epochs during which activity is significantly different from 0 (one-sample t-test, $p \leq 0.05$). Error bars / boxes represent s.e.m.

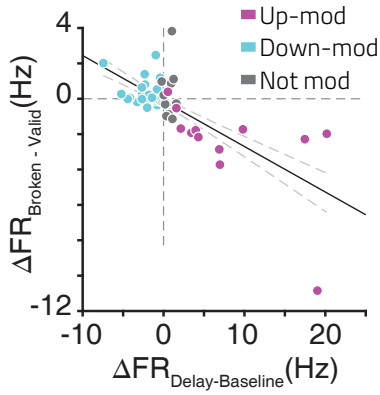


Figure 2.15. Unit engagement during the delay period is correlated with activity during broken fixation trials. Linear regression of the activity during the delay period ($FR_{Delay} - FR_{Baseline}$) and difference of activity aligned to broken fixations ([-0.1 : 0]s , as shown in Figure 2.14) for all photo-identified iMSNs (n = 89 units from 2 animals), labeled on the basis of their engagement during the delay period, relative to baseline (P $\ll$ 0.001, t_{87} = -9.43).

In contrast, iMSNs exhibiting decreased firing during action suppression, or not modulated, increased firing just before *broken fixations* ((Figure 2.14b,d)). Furthermore, the degree to which a cell was engaged during the delay period significantly correlated with differences of activity aligned to *broken fixations* (Figure 2.15). Thus, the biphasic photometric signal preceding *broken fixations* may reflect the contribution of two subpopulations of iMSNs. These data demonstrate that transient disruption of iMSN activity, as assessed both using photometry and electrophysiology, was associated with the failure to successfully suppress action.

2.2.3 Broadly opponent yet distinct functional contributions of striatal direct and indirect pathways to the control of action

To probe the functional importance of the observed patterns of neural activity in the two pathways, we next performed a series of optogenetic experiments. We combined the same Cre lines used to label MSNs in the previous experiments with cre-dependent viral expression of the light-activated proton pump ArchT(Han et al., 2011) or ChR2(Nagel et al., 2003) and implanted tapered optical fibers(Pisanello et al., 2017) in the dorso-lateral striatum to enable inhibition or activation of iMSN or dMSN activity(Fig. 3g-i). We first characterized the effect of photoinhibition on neuronal firing by performing extracellular electrophysiological recordings from striatal neurons during a quiet awake state in the absence of a behavioral task (Figure 2.16). Illumination of striatal tissue exclusively produced rapid and sustained inhibition of putative MSN firing in both iMSN-ArchT and dMSN-ArchT mice, indicating effective inhibition of the two pathways, with no evidence for disinhibition of non-targeted MSNs. In contrast, ChR2 mediated excitation produced robust activation of MSNs, but also sustained inhibition in 10-20% of putative MSNs. We thus proceeded to examine the behavioral effects of optogenetic inhibition during the task, as activation would appear to mix both activation and inhibition, potentially across the two MSN types, complicating the interpretability of activation experiments with respect to the role of specific cell types in the striatum. We next delivered light to the fiber(s) on a random minority of trials (30%), starting at trial initiation and ramping off over 250ms starting at either with the second tone onset or when *broken fixation* was detected, whichever occurred first (Figure 2.17). Bilateral optogenetic inhibition of iMSNs produced a near-complete inability of mice to suppress movement during interval presentation overall. Animals *broke fixation* on 35±6% (n=4 mice) of non-inhibited trials and on 86±7% of trials when iMSNs were inhibited bilaterally and were generally unable to maintain fixation for the longest interval duration. However, levels of *broken fixation* were unaffected by bilateral inhibition of dMSNs during equivalent trial epochs (Figure 2.18). On *broken fixation* trials followed by movement toward a choice port, the first 300ms

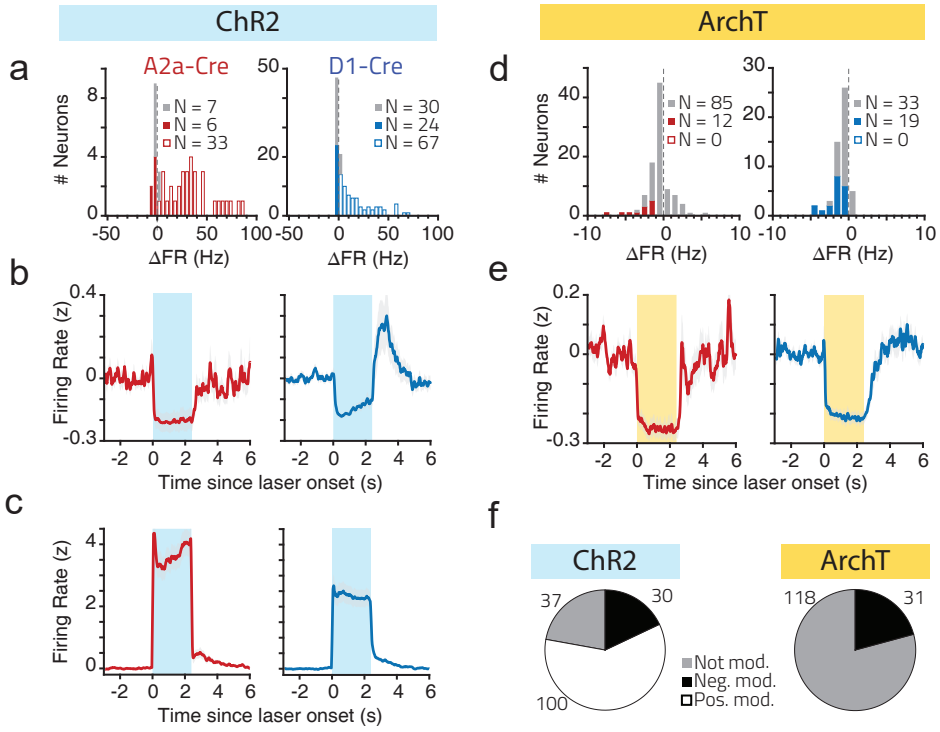


Figure 2.16. Acute characterization of the effects of optogenetic manipulation in putative medium spiny neurons. **a)** Distribution of changes in firing rate (Hz) during the period of light delivery, versus baseline, for putatively labeled MSN units recorded from A2a (Red) and D1-Cre (Blue) mice expressing ChR2, outside the context of the task. Gray depicts non-significantly modulated cells, closed and open shapes depict significantly down- and up-modulated cells, respectively. **b-c)** Overall average peristimulus time histogram (PSTH) of all negatively **(b)**, or positively **(c)** light-modulated cells, putatively labeled as MSNs, recorded from A2a-Cre and D1-Cre mice during the ChR2 acute experiment. All units were z-scored (see methods). Gray, open bars and closed bars depict not-modulated, negatively modulated and positively modulated units during the period of light delivery, respectively. Shaded area depicts the time of laser illumination. **d-e**). Same as **a-b)** but for animals expressing ArchT in medium spiny neurons. **f)** Summary of overall modulation effects of ArchT versus ChR2 activation in putative medium spiny neurons. Error bars represent s.e.m..

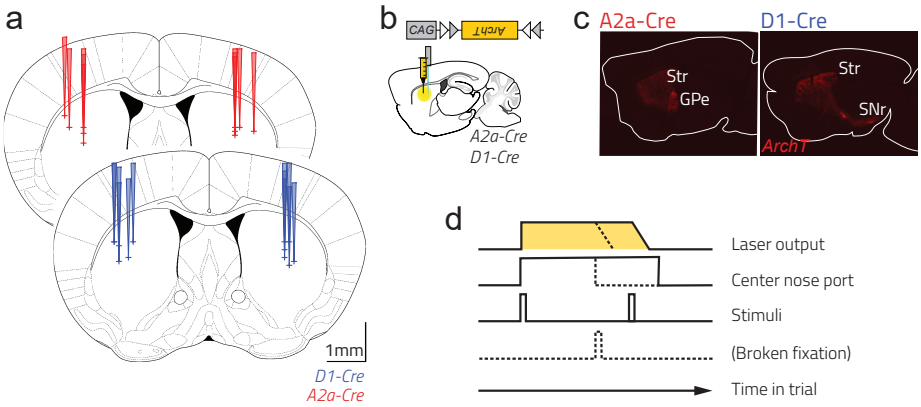


Figure 2.17. **a)** Histological reconstruction of sites of tapered fiber implantation used for ArchT optogenetics experiments in DLS. Animals are colored by their genotype according to the legend. All coordinates were projected to the same coronal slice (AP = +0.5 from bregma) adapted from Franklin & Paxinos (2008). **b)** Viral strategy and **c)** Pattern of transgene expression in A2a-Cre (iMSN) or D1-Cre (dMSN) animals in sagittal section, 2.1mm ML. Str-Striatum, GPe-Lateral globus pallidus, SNr-Substantia nigra pars reticulata. **d)** Protocol of optogenetic manipulation. Laser was turned on at trial onset and turned off at stimulus offset or broken fixation, whichever occurred first.

of this movement overlapped with the ramping off of the light stimulation (250ms ramp off plus 50ms grace period, see methods). We therefore asked whether these movements were affected by inhibition of either iMSNs or dMSNs. In contrast to the observed effect of iMSN inhibition on action suppression, we found that dMSN inhibition, but not iMSN inhibition, resulted in a significant increase in choice time during *broken fixation* trials (Figure 2.19), consistent with a role specifically for the direct pathway of DLS in augmenting the vigor of movements (Turner & Desmurget, 2010; Panigrahi et al., 2015; Dudman & Krakauer, 2016). Furthermore, DLS dMSN activity was specifically necessary for movement invigoration a brief time in advance of movement initiation, as choice times were not affected during valid trials, when the laser began ramping off during the time it took animals to initiate their choice movement, nor in a subset of experiments where inhibition was applied during execution of the choice movement, starting

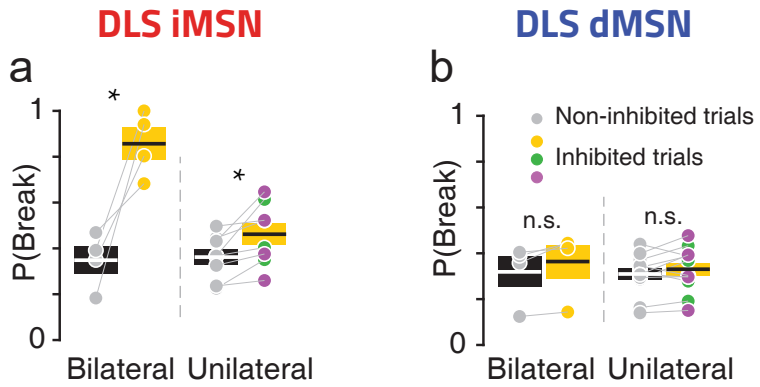


Figure 2.18. Overall probability of breaking fixation during iMSN **(a)** or dMSN **(b)** inhibition experiments. Colored and black dots represent data from laser-on and laser-off trials, respectively. "Bilateral" condition represents data from sessions wherein light was delivered bilaterally to the DLS (iMSN: one-sample t-test, P = 0.022, t_3 = 4.37, N = 4 pairs of hemispheres, dMSN: P = 0.074, t_3 = 2.689, N = 4 pairs of hemispheres) whereas "unilateral" represents data from sessions wherein light was delivered to a single hemisphere. (iMSN: P = 0.015, t_7 = 3.21, N = 8 hemispheres; dMSN: P = 0.185, t_{11} = 1.413, N = 12 hemispheres). Green/Purple code for CS/CL manipulation sessions, respectively). Error bars represent s.e.m.. n.s. P>0.05, *P≤0.05.

at the onset of the second tone (Figure 2.25). Alternatively, it is possible that cued (*i.e.* after second tone) and self-generated (*i.e.* *broken fixations*) actions asymmetrically rely on dorso-lateral striatal circuits, perhaps explaining why the vigor effect is circumscribed to the latter.

Given the observed interhemispheric dynamics in the photometry signals collected from DLS iMSNs (Figure 2.13) and behavioral indications of time varying lateralized motor plans that require suppression (Figure 2.12), we next asked whether unilateral iMSN inhibition would disrupt action suppression preferentially at the times when activity in a given hemisphere appeared to be most engaged. Indeed, while unilateral iMSN inhibition produced a more modest increase in *broken fixations* overall (Figure 2.18), the timing of *broken fixations* was systematically related to the laterality of inhibition and observed patterns of neural activity. Mice were more likely to *break fixation* early or late when

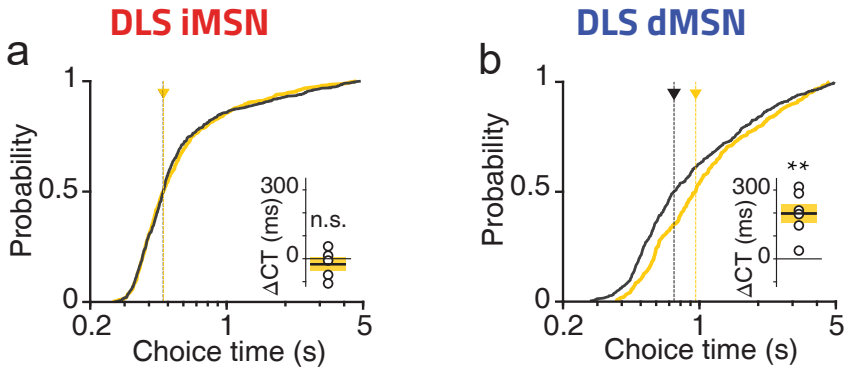


Figure 2.19. Cumulative distribution of choice times during iMSN **(a)** and dMSN **(b)** inhibition experiments (i.e. time to travel from the center port to a side port) after broken fixation trials for manipulated (yellow) and non-manipulated (black) conditions of all animals. Dashed lines show the median of the distributions across all animals. Inset shows the difference in medians of the two distributions for single animals (manipulated - non-manipulated, one-sample t-test, iMSN: P = 0.46, t_4 = -0.817, N = 5; dMSN: P = 0.005, t_5 = 4.834, n = 6 animals) Error bars represent s.e.m.. n.s. P>0.05, **P≤0.01.

iMSNs contralateral to the "short" or "long" choice were inhibited, respectively, as compared to non-inhibited trials (Figure 2.20a). Taking the difference in probability of *breaking fixation* between inhibited and control trials as a function of time during the trial gives a measure of the time-course of the effect on action suppression for each hemisphere. Superimposed, these two measures cross near the 1.5s decision boundary (Figure 2.20b,c), mirroring the precise contingency between reward and action over time during a trial, in striking correspondence with the pattern of hemispheric dominance observed in iMSN neural activity (Figure 2.13). We observed no consistent effects of unilaterally inhibiting dMSNs on the timing of *broken fixations*(Figure 2.21).

Lastly, given the widely assumed importance of BG circuits in action selection, we asked whether inhibition of MSNs affected the probability that particular actions were executed. When inhibiting iMSNs, we observed a consistent increase in the probability that an animal would execute a choice to the port contralateral to the site of iMSN inhibition, as compared to non-inhibition trials, after *breaking*

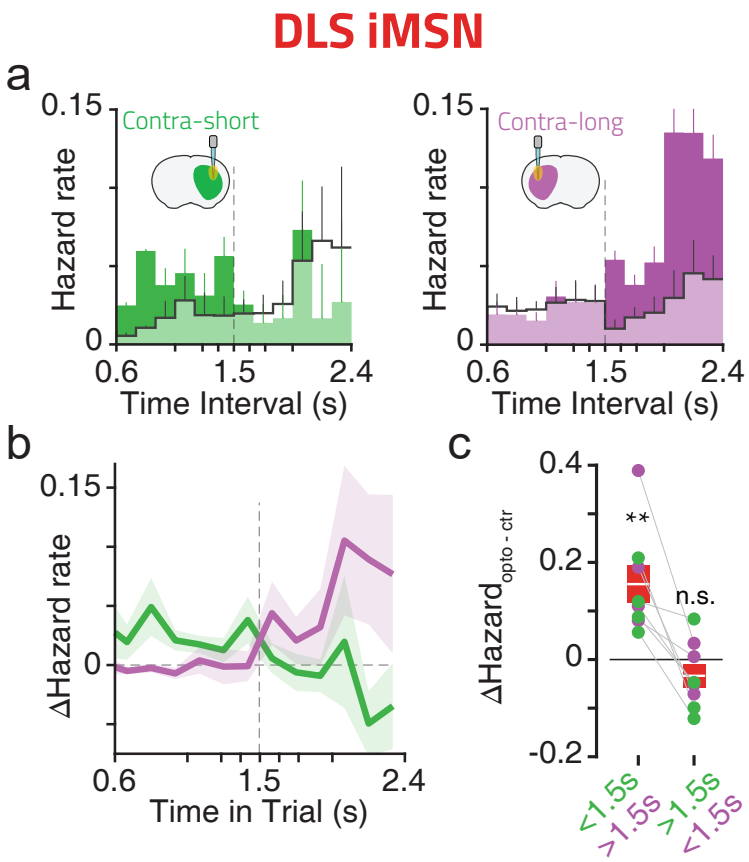


Figure 2.20. **a)** Hazard rate of breaking fixation over all included trials of a given condition, for control (black outline) and manipulated trials at the hemisphere contra-lateral to a short (green) or long (purple) choice port in A2a-Cre animals. **b)** Change in probability ($\Delta P = P_{Manipulation}$ - $P_{Control}$) due to inhibition of the CS (green) or CL (purple) relative to session matched control trials. **c)** Quantification of the effect shown in b). We calculated the hazard of breaking fixation during the period where the choice contralateral to the site of inhibition would be correct or incorrect (before/after 1.5s and after/before 1.5s for CS and CL, respectively). Data shown are the differences between session matched controls and manipulations. Each pair of points depicts data from the same hemisphere and the color the site of manipulation. (contralateral-correct: one-sample t-test, P = 0.005, t_7 = 4.055, contralateral-incorrect: P = 0.215, t_7 = -1.363, N = 8 Hemispheres). Error bars represent s.e.m.. n.s. P>0.05, **P≤0.01.

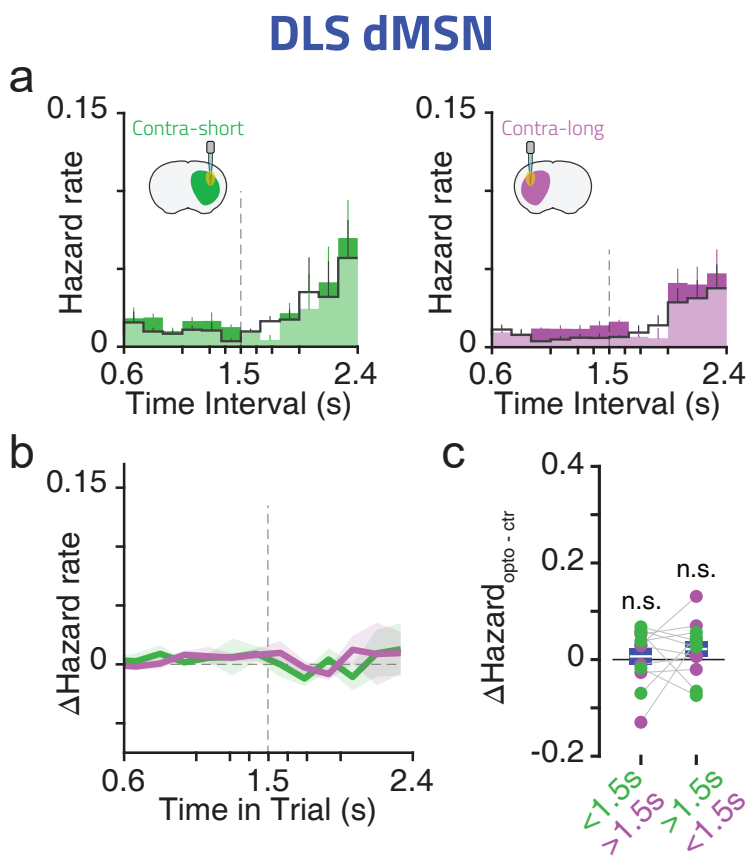


Figure 2.21. **a)** Hazard rate of breaking fixation over all included trials of a given condition, for control (black outline) and manipulated trials at the hemisphere contra-lateral to a short (green) or long (purple) choice port in D1-Cre animals. **b)** Change in probability ($\Delta P = P_{Manipulation}$ - $P_{Control}$) due to inhibition of the CS (green) or CL (purple) relative to session matched control trials. **c)** Quantification of the effect shown in b). We calculated the hazard of breaking fixation during the period where the choice contralateral to the site of inhibition would be correct or incorrect (before/after 1.5s and after/before 1.5s for CS and CL, respectively). Data shown are the differences between session matched controls and manipulations. Each pair of points depicts data from the same hemisphere and the color the site of manipulation. (contralateral-correct: one-sample t-test, P = 0.711, $t_{11} = 0.38$, contralateral-incorrect: P = 0.211, $t_{11} = 1.329$, N = 12 Hemispheres). Error bars represent s.e.m.. n.s. P>0.05.

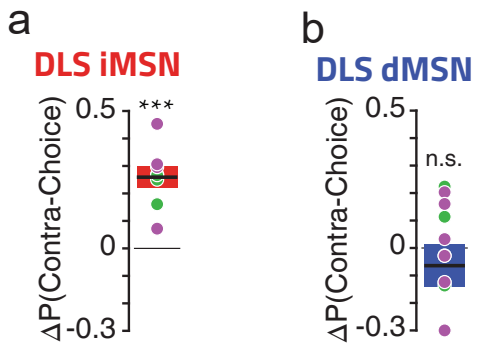


Figure 2.22. Change in probability of registering a choice at the port contralateral to the hemisphere manipulated, after breaking fixation ($\Delta P = P_{Manipulation} - P_{Control}$, in A2a-Cre (**a**), one-sample t-test, P $<$ 0.001, t_7 = 6.605, N = 8 hemispheres) or D1-Cre animals (**b**), P = 0.428, t_{11} = -0.824, N = 12 hemispheres). Error bars represent s.e.m.. n.s. P$>$0.05. ***P $\leq$ 0.001

fixation (Figure 2.22). This effect did not simply reflect the fact that animals were more likely to make short or long choices after early or late *broken fixations*, respectively, because it was present in *broken fixations* made both before and after the decision-boundary (Figure 2.23). Features of the choice movements following iMSN unilateral inhibition, as quantified through high-speed tracking of the animals' nape, revealed no consistent difference in trajectory as compared to choice movements performed in the absence of iMSN inhibition (Figure 2.24). Once again, we neither observed a significant effect of unilateral DLS dMSN inhibition on lateralized choice behavior after *broken fixations* (Figure 2.22), nor when dMSN inhibition was applied during execution of the choice movement (Figure 2.25). These data, together with the observed inter-hemispheric dynamics of the endogenous activity of iMSNs, demonstrate that the indirect pathway in DLS of a given hemisphere was dynamically engaged to suppress contralateral movements when those particular actions would be tempting and thus in greater need of suppression, and that such suppression of specific actions can contribute to determining whether a particular action is selected for execution.

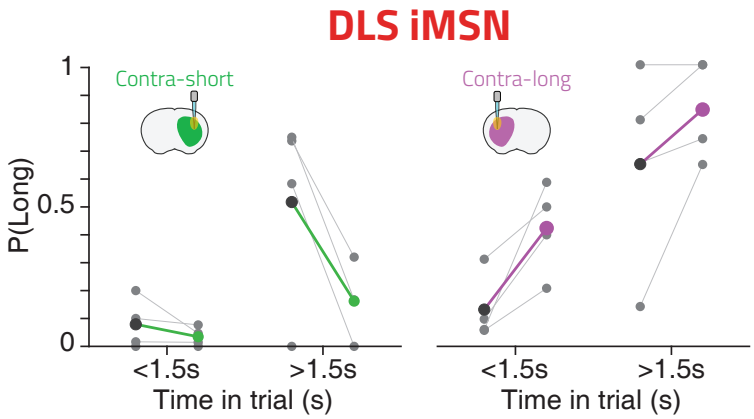


Figure 2.23. Bias to report a contra-lateral choice after inhibition of iMSNs is not explained by the tendency of mice to make particular choices after *breaking fixation* early or late in the delay. Each panel, one for manipulations performed in each hemisphere, depicts the data shown in Figure 2.22 further split by whether fixation was broken before or after the 1.5s decision boundary. Black dots represent control trials whereas color dots represent data from, session matched, manipulated trials.

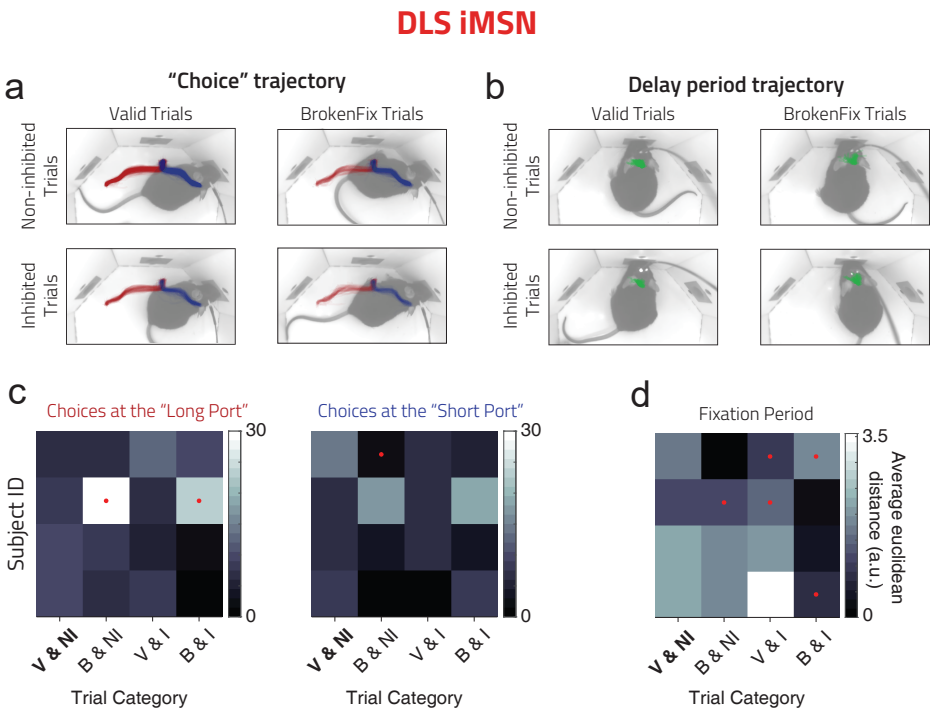


Figure 2.24. Unilateral DLS indirect pathway inhibition did not systematically affect movement trajectories across subjects. **a)** Example trajectories for a single animal aligned to center-out for choices to the "Long port" (red) or "Short pot" (blue) for completed trials (Left) and Broken fixation trials (Right). Trials are further broken down by whether the indirect pathway was inhibited (bottom) or not (top) in animals implanted in the DLS. **b)** same as a) but for trajectories measured during the delay period (up until second tone or broken fixation events). **c-d)** Quantitative differences between trajectory distributions among different conditions. Briefly, we computed a mean reference trajectory from "Valid & Non-inhibited" condition and computed, for each trial from each condition, the average euclidean distance to this reference trajectory. Values shown in the heatmaps correspond to the means of these distributions. Significance was accessed by computing a two-sample Kolmogorov-smirnov test between the reference and testing condition ($P \leq 0.05$ is reported as a red dots). c) and d) show the analysis performed during the choice (c) and fixation (d) epochs. Condition labels as follows: **V & NI**, Valid & Non-inhibited (Reference condition); **B & NI**, Broken fixation & Non-inhibited; **V & I**, Valid & Inhibited; **B & I**, Broken fixation & Inhibited;

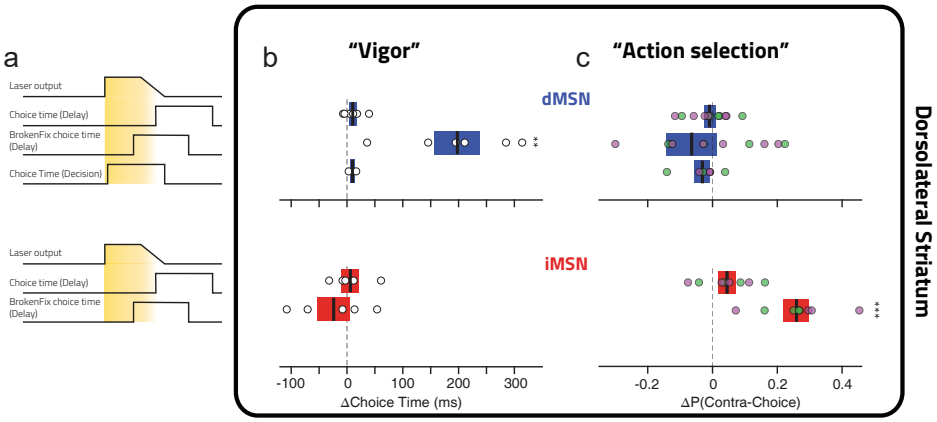


Figure 2.25. Summary of manipulation-induced changes in vigor and action selection in DLS optogenetics experiments. **a)** Cartoon depicts the three different manipulated trial types: Choice Time (Delay), laser was ramped off as the second tone is played. BrokenFix Choice Time (Delay), laser was ramped off as the animal leaves the centre port causing a broken fixation. MovementTime (Decision) laser was turned on as the second tone is played until the animal either performs its choice or 400ms elapse , whichever occurs first . **b)** Differences in single animal's median choice time between inhibited and non-inhibited trials ($\Delta ChoiceTime = ChoiceTime_{Manipulation} - ChoiceTime_{Control}$). For each animal, we concatenated all sessions and split trials in manipulated versus non-manipulated. From top to bottom: one-sample t-test, P = 0.198, t_5 = 1.482; P = 0.198, t_5 = 1.482;P = 0.005, t_5 = 4.834; P = 0.128, t_2 = 2.523; P = 0.718, t_4 = 0.388; P = 0.46, t_4 = -0.817; **c)** Differences in probability of reporting a contra-lateral choice, relative to inhibition side ($\Delta P = P(Contralateral\ choice)_{Manipulation} - P(Contralateral\ choice)_{Control}$. For each animal we concatenated trials from sessions with unilateral perturbation and normalized choices to the side contralateral to inhibition site. From top to bottom:P = 0.597, t_{11} = -0.545; P = 0.427, t_{11} = -0.824; P = 0.259, t_5 = -1.274; P = 0.149, t_7 = 1.623; P « 0.001, t_7 = 6.605; Boxes represent s.e.m..*P≤0.05, **P≤0.01, **P≤0.001

2.2.4 A simplified reinforcement learning model of action suppression

Based on the data presented thus far, DLS circuits appeared to be largely engaged to suppress tempting movements, suggesting that the drive to produce those movements was generated elsewhere. During learning, plasticity at striatal synapses is thought to underlie a computation similar to that of learning a policy, a program for behavioral control that consists of a mapping between states and actions within reinforcement learning (RL) models(Sutton et al., 1998; Doya, 1999). If the policy learned by DLS in this task is to suppress movement, how is this policy learned? Parallel BG circuits are known to integrate different types of information(Alexander & Crutcher, 1990), and are proposed to learn distinct types of policies Bornstein & Daw (2011), and thus we hypothesized that the suppressive policy of DLS circuits might arise in response to an action promoting policy located elsewhere in the striatum.

We thus constructed a simplified model containing two "sub-agents", whose policies were combined to perform the interval categorization task (see methods). Each sub-agent consisted of three main components: a state representation, a state value function, and action preference functions. The model followed an actor-critic architecture, where prediction errors generated according to a temporal difference algorithm were used to drive learning of both state values and action preferences (Figure 2.26). Our primary goal in constructing this model was to study how the patterns of activity of the direct and indirect pathway in DLS might be established. For this reason, one of the sub-agents, from here on referred to as the DLS sub-agent, was endowed with a direct and indirect pathway. To account for known differences in how dopaminergic prediction errors affect plasticity mechanisms at dMSN and iMSN synapses(Shen et al., 2008; Collins & Frank, 2014; K. N. Gurney et al., 2015; Iino et al., 2020; S. J. Lee et al., 2021), the direct and indirect pathways of the DLS sub-agent were subject to different learning rules that were sigmoid functions of prediction error (Figure 2.27). Action suppression signals in the indirect pathway were strengthened by negative prediction errors, whereas action preference signals in the direct pathway were strengthened by positive prediction errors. For simplicity, the second sub-

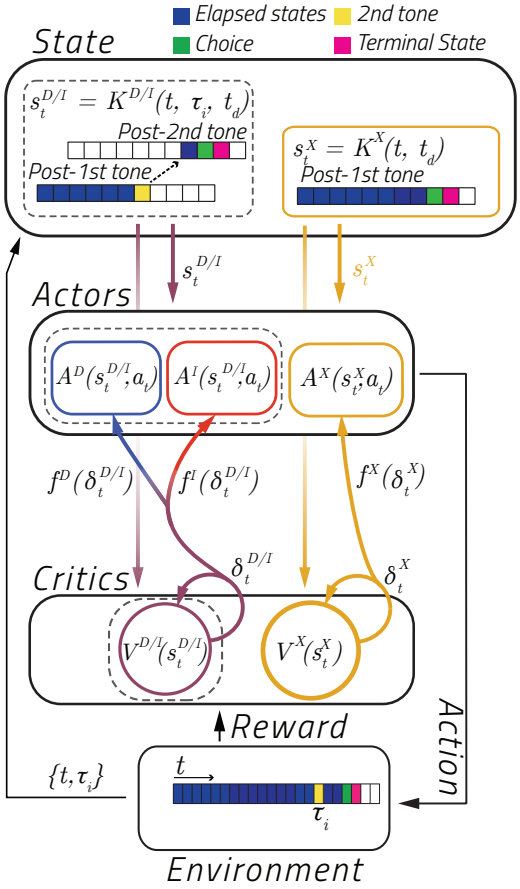


Figure 2.26. Schematic of the multi-agent actor-critic reinforcement learning model. The model consists of two critics $\{V^{D/I}(s_t^{D/I}), V^X(s_t^X)\}$, one with access to a state representation $s_t^{D/I}$ given by $K^{D/I}(\theta, \Delta\theta_i^d, \tau_i)$ which contains post-second tone states and the other with a state representation $K^{D/I}(\theta, \Delta\theta_i^d)$ which only contains pre-second tone states. After receiving a reward R each critic updates its value and connects to a set of corresponding actors given by the action preference values $\{A^D(s_t^{D/I}, a_t), A^I(s_t^{D/I}, a_t), A^X(s_t^X, a_t)\}$ each receiving a filtered version of the respective generated Reward Prediction Errors (RPEs) $\{\delta_t^{D/I}, \delta_t^X\}$ that correspond to the DLS agents mapping onto dMSNs $f^D(\delta_t^{D/I})$, iMSNs $f^I(\delta_t^{D/I})$, and the non-DLS agent $f^X(\delta_t^X)$ (Figure 2.27). The actors are then combined to generate actions, taking into account the relative positive and negative weights of each pathway.

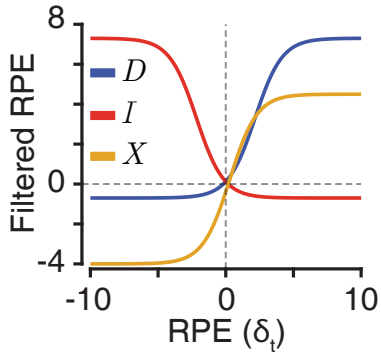


Figure 2.27. Functions used to generate actor-specific RPEs. Blue: DLS dMSNs, $f^D(\delta_t^{D/I})$; Red: DLS iMSNs, $f^I(\delta_t^{D/I})$; Gold and the non-DLS agent $f^X(\delta_t^X)$

agent was modeled in a more standard fashion as possessing a single pathway. Critically, the two sub-agent modules were provided with different state representations. The DLS sub-agent received information that more closely resembled the immediate state of the environment, encoding delay from the first tone as well as information about the occurrence of the second tone that defined interval offset. The other agent received a more abstract representation suitable for learning whether the world was in a "short" or "long" state occurring before or after 1.5s since first tone delivery. Though not explicitly modelled as such, this representation can also be viewed as predictive(Dayan, 1993a; Stachenfeld et al., 2017), encoding information about the states subsequent to the occurrence of a second tone were it to occur in that moment. Each sub-agent learned a set of action preferences from its experience in the task, which were then combined to guide action selection (see methods).

With this architecture, despite the two sub-agents learning very different policies for behavioral control, the model was able to learn to make appropriate choices, producing psychometric curves similar to those of mice (Figure 2.28 a). Also as observed in the mice, the model sometimes produced premature choices that were nonetheless directed towards the appropriate sides as a function of time (Figure 2.28 b,c). These *broken fixations* were the consequence of the more compact state representation received by the single pathway sub-agent, which

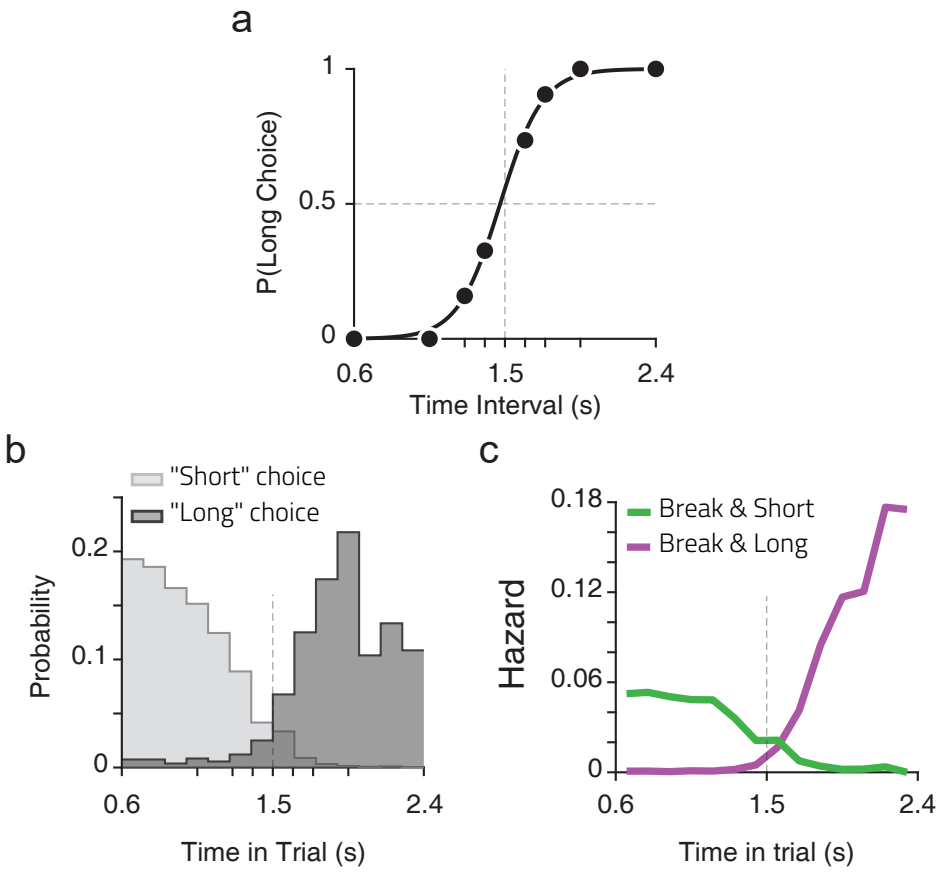


Figure 2.28. **a)** Psychometric fit (black line) to the performance of the model for each interval (black dots) used in the time interval categorization task. **b)** Probability density functions of broken fixations, produced by the model agent, over time (0.6s to 2.4s), conditioned on subsequent choice at one of the side ports (light grey: *SHORT* choice, dark grey: *LONG* choice). **c)** Hazard rate of broken fixation conditioned on subsequent choice at one of the side ports (*SHORT* and *LONG* as green and purple, respectively).

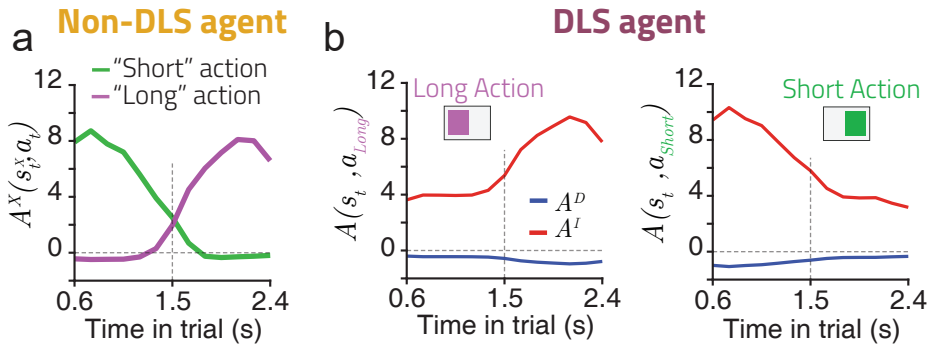


Figure 2.29. Action preference values, $A(s, a)$, for the two actions: *LONG* and *SHORT*, modeling contra-long and contra-short hemispheres, respectively. a) Action preferences relative to the non-DLS agent, $A^X(s_t^X, a_t)$ for the *LONG* (purple) and *SHORT* (green). **b)** Action preferences for two actors ($A^D(s_t^{D/I}, a_t)$, in blue and $A^I(s_t^{D/I}, a_t)$, in red) that make up the DLS agent. Left and right show the action preferences for the *LONG*, $A(s_t, a_{LONG})$, and *SHORT*, $A(s_t, a_{SHORT})$, actions, respectively.

learned to exert a drive to make short choices before 1.5s, and long choices after, constituting an urge to act (Figure 2.29 a). The DLS sub-agent learned to suppress this urge until the second tone was delivered due to the negative RPEs produced by premature choices. These negative RPEs drove the development of increased suppressive signals in the indirect pathway when the contralateral choice action was tempting (Figure 2.29 b). The model thus reproduced multiple features of mouse behavior as well as the qualitative patterns of overall and interhemispheric activity observed in the DLS of mice performing the task. We next modeled the effect of optogenetic inhibition of DLS iMSNs by selectively decrementing the indirect pathway of the DLS sub-agent. As in mice, inhibition of the indirect pathway led to pronounced failure of the model to suppress action (Figure 2.30 a). Furthermore, inhibiting the model's indirect pathway "short" action preference function led to failed action suppression early in the delay period, whereas inhibiting the indirect pathway "long" action preference led to failure to suppress action late in the delay period (Figure 2.30 b), recapitulating the results of unilateral iMSN inhibition in mice. In addition, unilateral inhibition of the indirect pathway caused an increase in the probability of contralateral action

(Figure 2.30 c), as in the experimental data (Figure 2.22). Previous data have shown that optogenetic activation of DMS MSNs can affect action selection(Tai et al., 2012). We thus wondered whether more medial circuits in the DMS might contribute to the action promoting functions reflected in the single pathway sub-agent of the model. We repeated the optogenetic inhibition experiments carried out in DLS, but now targeting the DMS in separate groups of mice. In general, behavioral effects of inhibiting DMS MSNs were more modest in magnitude than those observed when inhibiting DLS MSNs and we observed no significant effect of inhibiting either iMSNs or dMSNs in DMS on action suppression (Figure 2.25, Figure 2.32). However, we did observe a modest but significant decrease in the probability of contralateral choice after *broken fixations* when inhibiting DMS dMSNs. This effect was opposite in sign to that observed when inhibiting iMSNs in DLS (Figure 2.22), and could be replicated in the model by unilaterally inhibiting the single pathway sub-agent (Figure 2.30 c). These data suggest that more medial circuits contributed to an opposing function to that of DLS, promoting specific movements when they would potentially be rewarded through, at least in part, the action of dMSNs.

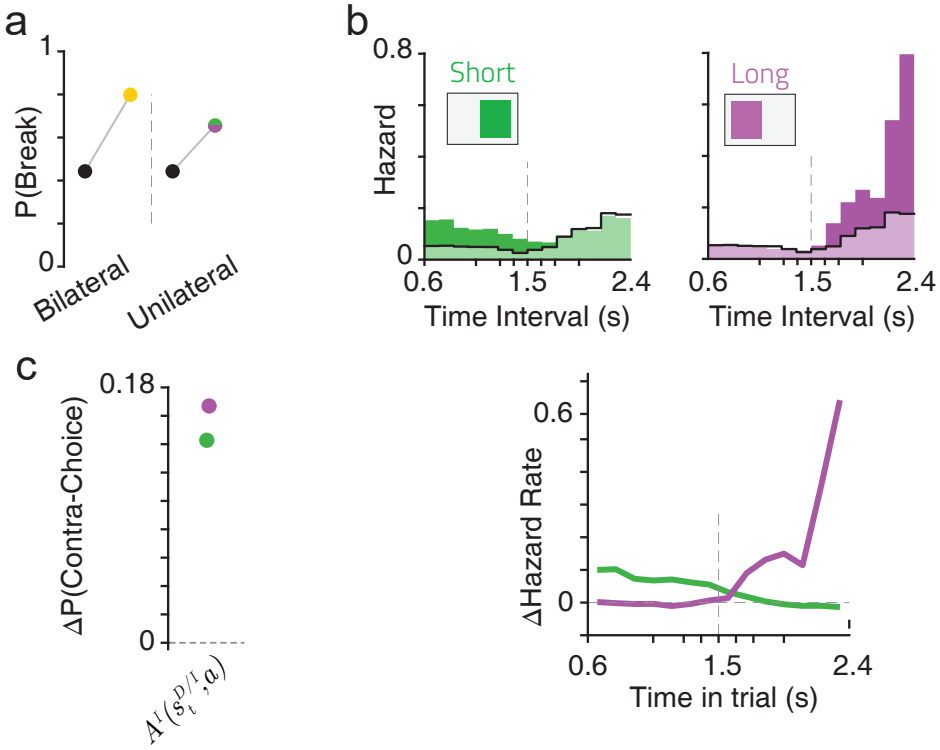


Figure 2.30. Reducing the action preference values of the DLS indirect pathway in the model ,$A^I(s_t^{D/I}, a_t)$, leads to qualitatively similar results to those observed in the optogenetic inhibition experiments. **a)** Effect of reducing action preference values for the two actions simultaneously ('Bilateral', yellow) or for only SHORT or LONG ('unilateral' green/purple) on the probability of breaking fixation. **b)** Top: Effect of reducing action preference values on the hazard rate of breaking fixation over all included trials of a given condition, for control (black outline) and manipulated trials on the *SHORT*(green) or *LONG*(purple) actions. Bottom: Change in probability ($\Delta P = P_{Manipulation}$ - $P_{Control}$) of breaking fixation from the data depicted on top. **c)** Change in probability ($\Delta P = P_{Manipulation}$ - $P_{Control}$) of generating a contralateral action when decreasing the magnitude of action preference values of the indirect pathway.

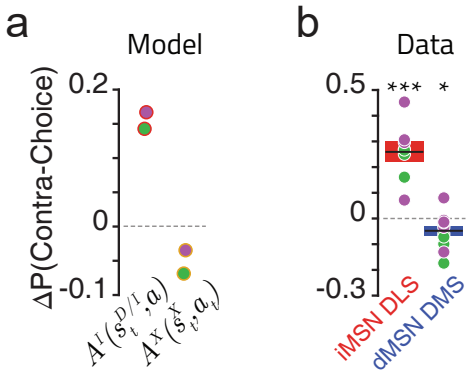


Figure 2.31. **a)** Change in probability of generating a contralateral action when decreasing the magnitude of action preference values, of the indirect pathway,$A^I(s_t^{D/I}, a_t)$ (red outline), or pathway X, $A^X(s_t^X, a_t)$ (gold outline), for one of the two actions (SHORT or LONG) before the second tone observation (i.e. broken fixations). **b)** Change in probability of registering a choice at the port contralateral to the hemisphere manipulated when inhibiting iMSNs in DLS (red, same data as in Figure 2.30) or dMSNs in DMS (blue, one-sample t-test, P = 0.028, t_{11} = -2.528, N = 12 hemispheres), after breaking fixation.

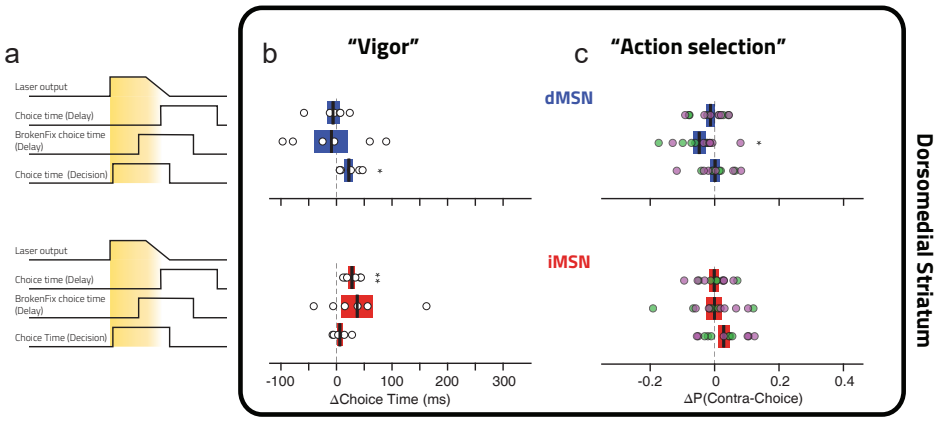


Figure 2.32. Summary of manipulation-induced changes in vigor and action selection in DMS optogenetics experiments. **a)** Cartoon depicts the three different manipulated trial types: Choice Time (Delay), laser was ramped off as the second tone is played. BrokenFix Choice Time (Delay), laser was ramped off as the animal leaves the centre port causing a broken fixation. MovementTime (Decision) laser was turned on as the second tone is played until the animal either performs its choice or 400ms elapse , whichever occurs first . **b)** Differences in single animal's median choice time between inhibited and non-inhibited trials ($\Delta ChoiceTime = ChoiceTime_{Manipulation} - ChoiceTime_{Control}$). For each animal, we concatenated all sessions and split trials in manipulated versus non-manipulated. From top to bottom: one-sample t-test, P = 0.617, t_5 = -0.533; P = 0.774, t_5 = -0.304; P = 0.032, t_5 = 2.959; P = 0.005, t_5 = 4.876; P = 0.247, t_5 = 1.309; P = 0.306, t_5 = 1.141. **c)** Differences in probability of reporting a contra-lateral choice, relative to inhibition side ($\Delta P = P(Contralateral\ choice)_{Manipulation} - P(Contralateral\ choice)_{Control}$. For each animal we concatenated trials from sessions with unilateral perturbation and normalized choices to the side contralateral to inhibition site. From top to bottom:P = 0.365, t_{11} = 1.623; P = 0.028, t_{11} = -2.528; P = 0.94, t_{11} = 0.077; P = 0.933, t_{11} = -0.086; P = 0.985, t_{11} = -0.02; P = 0.143, t_{11} = 1.577; Boxes represent s.e.m..*P≤0.05, **P≤0.01, **P≤0.001

2.3 Discussion

Here we demonstrate clear activity signatures of large-scale functional opponency between neurons initiating the two major BG pathways in a normal, non-pathological state. In particular, we observe opposite patterns of activity in the two pathways in sensorimotor striatum when movements must be proactively and persistently suppressed.

Action suppression can be broadly separated into two classes. Reactive suppression involves stopping behavior in course when presented with an external stimulus. In contrast, proactive suppression involves selectively inhibiting particular response tendencies using advance knowledge. The behavioral context we studied requires proactive suppression of time varying response tendencies to move to the left or right. When subjects were required to suppress the urge to move in a given direction, on average iMSNs in DLS located contralaterally to that direction exhibited higher levels of activity than iMSNs located ipsilaterally. Consistent with these data, functional magnetic resonance imaging data in humans and electrophysiological data in non-human primates suggests that iMSNs might be selectively engaged to proactively suppress action (Majid et al., 2013; Watanabe & Munoz, 2010; Ford & Everling, 2009; Amita & Hikosaka, 2019). Leveraging the genetic access to iMSNs and dMSNs afforded by the use of mice as a model organism, the current data establishes not only that iMSNs in the DLS were selectively engaged as animals suppressed particular lateralized response tendencies, but that iMSNs, and not dMSNs, were necessary for it. In addition, successful action suppression relied on iMSNs in a given hemisphere at different points in time depending on learned task demands, indicating that specific subpopulations of iMSNs can be deployed to dynamically shape action suppression in time. While multiple studies to date have observed coactivation of dMSNs and iMSNs around movement(Cui et al., 2013; Tecuapetla et al., 2014; Markowitz et al., 2018), a finding we reproduce here (Figure 2.33), transient decorrelation of activity between the two pathways has been reported around transitions between actions in a behavioral sequence(Markowitz et al., 2018). Given our observations that activity in the two pathways is decorrelated or even anti-correlated during proactive action suppression, we hypothesize that previously reported transient

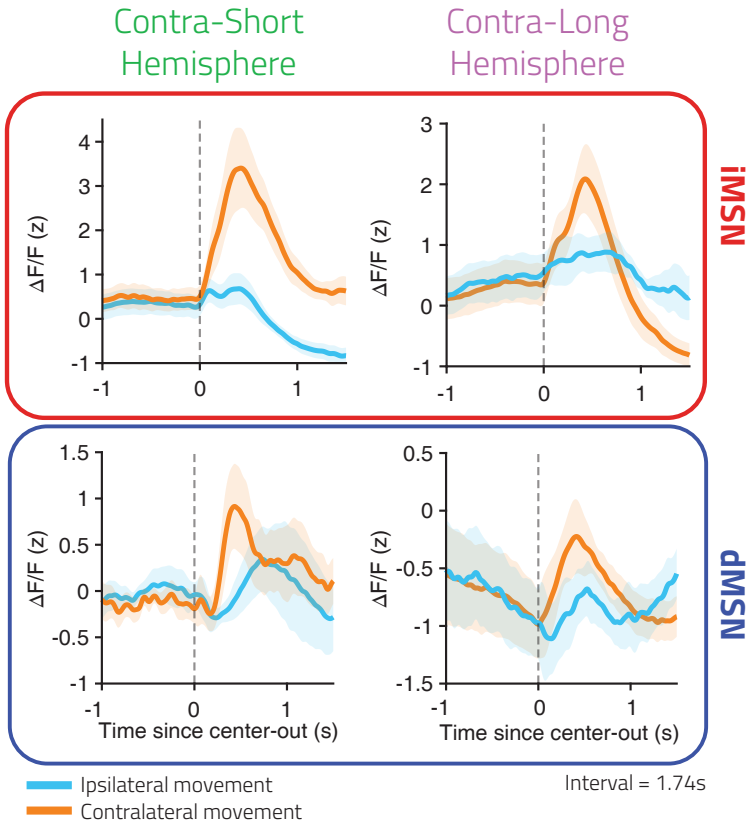


Figure 2.33. Both DLS direct and indirect pathways are more active during contra-lateral movements. Photometry signal aligned to leaving the center-port during a near boundary interval (1.74s) where the same stimuli results in different choices. Same dataset Figure 2.9. Error bars represent s.e.m.

decorrelations between the two pathways may arise when the brain must largely suppress the production of actions at behavioral transitions. The current data also demonstrate distinct contributions of dMSNs and iMSNs in the DLS to aspects of motor function beyond proactive action suppression. Bradykinesia seen in PD patients is thought to result primarily from the loss of dopamine neurons in the substantia nigra pars compacta, particularly those targeting sensorimotor striatum(Albin et al., 1989). Prior studies have identified that both dopaminergic

input to the striatum and the activity of striatal neurons, in particular dMSNs, are important for invigorating movement(Panigrahi et al., 2015). However, it was not clear based on previous work whether iMSN activity is necessary for invigorating movement. Here we show that inhibiting DLS dMSN activity slowed movement, consistent with previous work, without affecting which action was selected. Conversely, we show that iMSN inhibition did not lead to less vigorous movements, but instead disrupted the suppression of actions, and that this disruption reflected at least some degree of action specificity in iMSN function as it occurred alongside a change in the likelihood that lateralized actions were produced. Based on the current data, it remains uncertain the degree to which iMSNs act on specific actions, although in sensorimotor striatum we might expect iMSNs in a given location to act on actions involving parts of the body represented in the somatosensory and motor cortical areas that provide input to that striatal location(Hintiryan et al., 2016).

Interestingly, recovery of function in an experimental model of PD has been found to be associated with a return to near normal levels of iMSN but not dMSN activity, consistent with some degree of primacy for the sensorimotor indirect pathway in action production(Parker et al., 2018). Together with the data presented here, these results suggest the intriguing possibility that the indirect pathway can be configured locally in sensorimotor striatum to provide an inhibitory "mask", specifying which actions not to produce, while the the sensorimotor direct pathway provides a gain signal, as opposed to an action selection signal(Mink, 1996), on commands that are pushed through the mask. In this view, the selective function of the sensorimotor BG circuits on actions can be predominantly produced through the indirect pathway, a novel proposal that should serve as a basis for future experiments. In the present data, both pathways were more active around movements toward the contralateral side of the recording site (Figure 2.33), suggesting that suppressing and promoting signals around movement may be targeted towards nearby regions in the space of possible actions, perhaps in part by mechanisms in other brain systems such as the cortex, thalamus, and cerebellum (Mink, 1996; Park et al., 2020), or through reentrant circuits embedded in other striatal regions.

While previous work has identified correlations between MSN firing and kinematic or motivational variables(Lau & Glimcher, 2007; Rueda-Orozco & Robbe, 2015; Markowitz et al., 2018; Klaus et al., 2017), the lack of conclusive information regarding the tuning properties of individual MSNs has been a recurring issue in efforts to determine the functional importance of the direct and indirect pathways. We were able to circumvent the general problem of diverse and unknown selectivity of neurons in two ways. First, by training animals to remain immobile for an extended period in the task, we pushed the brain toward a state of suppressing action, a behavioral manipulation we hypothesized to have a common effect on many neurons in a given cell class regardless of their selectivity for actions. Second, the striatum in a given hemisphere shows enhanced functional involvement in contralateral movements(Schwarting & Huston, 1996; Kravitz et al., 2010). By training animals to perform a task wherein the relative value of left/right lateralized movements was varied over time during a prolonged period of action suppression, and observing or manipulating activity on one hemisphere at a time, we demonstrate how dorso-lateral striatal circuits of the BG can be configured to control specific elements of action suppression and production.

Lastly, we present a simple computational model that provides a parsimonious explanation of how patterns of cell type specific activity arise in DLS in relation to their revealed functional roles, as well as apparently opponent roles between more associative and more sensorimotor circuits. It has long been appreciated that the BG is characterized by a parallel circuit architecture, carrying sensorimotor, associative, and limbic information from distinct sets of outlying functional territories through the main axis of its circuitry in a largely segregated manner(Alexander et al., 1986; Alexander & Crutcher, 1990; Prescott et al., 2006; Lau & Glimcher, 2007) . While different functions have been proposed for the different parallel circuits, emphasis within theoretical accounts of circuit function is usually placed on the need to arbitrate between types of control depending on context or task demands, for example between model-based and model free(Daw et al., 2005a), or pavlovian versus instrumental control(Dorfman & Gershman, 2019). Our model highlights that, in addition to being subject to arbitration, parallel circuits may learn distinct control signals that interact with each other

in their development, depending not just on context or task demands, but on the control signals being learned in other parallel circuits. This creates a situation within the brain akin to a multi-agent RL scenario where the policy of individual agents depends not only on the environment, but on the policies of the other agents in that environment. Once there are multiple, parallel drives on behavior, many of which may be anticipatory, and potentially conflicting with each other(Dayan et al., 2006), the system may require a mechanism to suppress in addition to promote action. We note that such a parallel, multi-process view of BG function may provide a natural explanation for the evolution and existence of a suppressive indirect pathway, linking RL based control to more classic proposals regarding BG circuit function (Mink, 1996).

Adaptive behavior fundamentally involves the interplay of action promoting and action suppressing mechanisms. The data presented here demonstrate that in sensorimotor striatum, elements of the direct and indirect BG pathways can express opposite patterns of modulation and be required for generally opponent yet distinct promoting and suppressing aspects of motor function. Knowledge of how circuits in other regions of the striatum, the BG at large, or elsewhere in the brain mediate this interplay represents a critical avenue toward a fundamental understanding of animal behavior. Such knowledge also has the potential to inform the engineering of artificial systems that can behave appropriately in complex environments, as well as how to design effective therapies for neurological and neuropsychiatric disease.

2.4 Methods

2.4.1 Animals

Adult (over 2 months) male and female mice of A2a:cre (KG139) (Gerfen et al., 2013), D1:cre (EY217)(Gerfen et al., 2013) and Ai32 (RCL-ChR2(H134R)/EYFP) (Madisen et al., 2012) lines were used for this study under the protocol approved by the Champalimaud Foundation Animal Welfare Body (Protocol Number: 2017/013), the Portuguese Veterinary General Board (Direcção-Geral de Veterinária, project approval 0421/000/000/2018) and in accordance with the European Union Directive 2010/63/EEC. Mice were group housed prior to surgical procedures and singly housed following surgery in an inverted 12h dark/light cycle. Mice were maintained under water deprivation for all behavioral experiments (>80% body weight from baseline ad libitum period before deprivation).

2.4.2 Behavioral apparatus

The behavioral box (20 x 17 x 19 cm), contained 3 nose ports and a speaker. The behavioral box consisted of 3 front walls (135 degree angle between the center and the side walls) 2 side walls and a back wall with a 90 degree angle between them. Each of the three front walls contained a nose port equipped with an infrared emitter/sensor pair to access port entry and exit times. The central nose port was defined as the trial initiation port, and choices were reported at the lateral nose ports. For correct trials, a 4-6 µL calibrated water reward was delivered using a solenoid valve. Tones were delivered through a speaker mounted on the center wall. All sensors and effectors in the behavioral box were read and controlled using a microprocessor (Arduino Mega 2560, Arduino) via a custom circuit board. The task was implemented by the microprocessor, which outputted data via a serial communication port to a desktop computer running custom Python-based software. High-speed video was acquired at 120fps and 640*480 pixel resolution (FL3-U3-13S2, FLIR) using Bonsai(Lopes et al., 2015).

2.4.3 Behavioral Task

Mice were trained to categorize interval durations as either short or long by performing right and left choices as previously described in Soares et al. (2016). Briefly, mice self-initiated trials by entering the central nose port, triggering the delivery of a pair of tones (7,500 Hz, 150 ms) separated by one of 6 randomly uniformly sampled selected intervals (0.6, 1.05, 1.26, 1.74, 1.95 and 2.4 s or 0.6, 1.26, 1.38, 1.62, 1.74, and 2.4s). Stable performance was usually achieved after 3-4 months of training. Trial availability was not signaled to the animal but inter-trial onset interval was kept constant within each animal (7-8 s). Thus, initiation port entries before the point that a trial became available were ineffectual. After the first tone was presented, mice were required to maintain interruption of the center nose port IR beam until the second tone was delivered; we refer to this action as "fixation" throughout the text. If the mouse departed the port before the second tone, an error tone (150ms of white noise) was played and the next trial availability delayed (timeout). We refer to these trials as *broken fixations*. To prevent incorrectly flagging trials as *broken fixations* due to sporadic state transitions in the IR beam, we only counted a trial as broken fixation after the beam had been continuously uninterrupted for 50 ms. After both tones were played, mice reported their judgments by entering one of the two laterally located nose ports over the next 10 seconds. For intervals shorter than a 1.5 s category boundary, responses were reinforced at one of the lateral ports. For intervals longer than 1.5 s, responses were reinforced at the opposite port. Incorrect responses were followed by a white noise burst (150 ms) and a timeout (12-18s inter-trial onset interval). The short/long vs right/left contingencies were counterbalanced across animals. Therefore, we adopt the nomenclature of Contra-short/Contra-long (or Ipsi-long/Ipsi-short, respectively) hemisphere throughout the paper. Sessions typically lasted 2 hours.

Psychometric functions were fit using a 4-parameter logistic function:

$$P(x) = (u - l) \cdot \frac{e^{\frac{x-b}{s}}}{1 + e^{\frac{x-b}{s}}} + l \tag{2.1}$$

where P is the performance of the animal on interval x, u and l the upper and lower asymptote of the curve, respectively, b the bias parameter and s the slope parameter.

We excluded broken fixation trials before 100ms, as these often reflected the failure of animals to properly settle in the initiation port and rarely resulted in a subsequent choice on one of the side ports.

To calculate hazard $H(k)$ of breaking fixation at a particular time bin k (defined between t and $t + \Delta t$), we used the following equation:

$$H(k) = \frac{B(k)}{\sum_{j=k}^{T} B(j) + \sum_{j=k}^{T} C(j)}, t < k < t + \Delta t \tag{2.2}$$

where $B(k)$ is the number of broken fixation trials that occured between t and $t + \Delta t$, $\sum_{j=k}^{T} B(j)$ the sum of all *broken fixations* that occured in the time greater or equal to t, up until the longest possible interval T (2.4 seconds), and $\sum_{j=k}^{T} C(j)$ the sum of all completed trials that occured from time t until T.

2.4.4 Viral injections and fiber implantation

All surgeries were performed with mice under isoflurane anesthesia (1-2After achieving stable performance (usually >3 months), mice were allowed to regain baseline weight and were subject to viral injection and fiber implantation in the same surgery. Mice health was assessed daily and after at least 5 days, the water deprivation regime was reinstated. To make sure subjects recovered their pre-surgery performance before data collection sessions (photometry or optogenetics), they were gradually retrained in the task without and then with fibers attached. Upon reaching stable performance (1-2 weeks), data collection began. We alternated recorded/manipulated hemispheres every day. For fiber photometry experiments, we injected 300nL of a mixture of two viruses: AAV1-Syn-Flex-GCaMP6f (titer 1 x $10^1$3 gc/mL; University of Pennsylvania Vector Core) and AAV1-CAG-Flex-tdTomato (titer 0.5 x 10^{12} gc/mL; University of Pennsylvania Vector Core), at a 5:1 ratio, in DLS striatum (single injection, AP 0.5mm, ML

2.1mm, DV 2.6mm from pia) using an automated microprocessor controlled microinjection pipette with micropipettes pulled from borosilicate capillaries (Nanoject II, Drummond Scientific). Injections were performed at 0.2 Hz with 2.3 nL injection volumes per pulse. For all injections, the micropipette was kept at the injection site 10 min before withdrawal. After injection, we implanted, bilaterally, 2 fibres (MFC_200/245-0.53_ZF1.25(G)_FLT, DORIC LENSES) $200\mu m$ above the injection site. For optogenetic inhibition experiments, we injected AAV5-CAG-FLEX-ArchT-tdTomato (titer 10^{13} gc/mL, Addgene) in the DLS. For each hemisphere, we made two injections (500nL, AP 0.5mm, ML 2.1mm, DV 2.7mm and 3.1mm from bregma) and implanted one fibre (Lambda-B fibre 200μm core, 0.39NA, 1.5mm emitting length, Optogenix) near the injection site (AP 0.5mm, ML 2.1mm, DV 3.5mm from bregma). We used an identical protocol for DMS inhibition but targeting AP 0.7mm, ML 1.3mm DV 3.4mm.

2.4.5 Fiber photometry

The photometry apparatus was adapted from Matias et al. (2017). For all experiments, a single blue laser was coupled to a patchcord (100 μm core diameter, 0.22 NA) and connected to a collimator adapter (EFL 4.5 mm, NA 0.50) and a neutral density filter. Dichroic mirrors were fixed inside the main unit, allowing for 473 nm light delivery and GCaMP6f and tdTomato fluorescence detection. The 473 nm light was coupled into a patchcord (200 μm core diameter, 0.48 NA) using a lens (EFL 4.5 mm, NA 0.50) and a rotatory joint. The patchcord was mated to one of two chronically implanted optical fibers (200 μm core diameter, 0.48 NA). Laser intensities at the patchcord tip, before mating to the chronically-implanted fiber, were 15-40μW. For detection of GCaMP6f fluorescence, light was collected by the lens, transmitted and reflected by the dichroics before final filtering and focusing into a photodetector. For detection of tdTomato fluorescence, light was collected by the lens and transmitted through all dichroics before final filtering and focusing into a second photodetector. Photodetector output was digitized at 1 kHz (PCIe 6351, National Instruments) and recorded using custom software in Bonsai.

2.4.6 Fiber photometry data analysis

All photometry data analysis was performed with custom MATLAB software. Raw data was downsampled to 100Hz and low-pass filtered at 20Hz. Slow fluctuations were removed by subtracting a fitted polynomial to the raw signal (order < 5). For each session, $\Delta F/F$ was calculated for both channels as $\Delta F/F_t = (F_t - F_0)/F_0$, where F_0 was calculated as the 10th lower percentile from the filtered signal. Similarly to(Soares et al. 2016), robust regression using GCaMP6f and tdTomato $\Delta F/F$ was performed and the coefficient estimates were used to calculate a predicted GCaMP6f $\Delta F/F$ based on the observed tdTomato $\Delta F/F$. This predicted GCaMP6f $\Delta F/F$ was then subtracted to the observed GCaMP6f $\Delta F/F$F to calculate the corrected $\Delta F/F$. In order to compare across sessions, each session's corrected $\Delta F/F$ was z-scored using the mean and standard deviation calculated from a baseline period (5 to 2 seconds before trial onset). Signals for individual trials were then re-zeroed by subtracting the average of a period of 5 to 2 seconds before trial onset. In order to compare the rate of change of activity aligned to *broken fixations* to time-matched valid trials (Figure 2.14) we began by calculating the first derivative of the corrected $\Delta F/F$ for each trial. For each broken fixation trial, we aligned the trace to the timestamp of the broken fixation and, in order to compare to a time-matched valid trial, we cropped the average trace of all valid trials to the same length and aligned to the time of the broken fixation. We repeated this process for all broken fixation trials, yielding a broken fixation and respective control, time-matched dataset of trials. The steps of the analysis are graphically represented in Figure 2.34.

2.4.7 Acute recordings & analysis

To confirm the ability to inhibit medium spiny neurons, we performed acute recordings in the dorsal striatum of untrained animals. Briefly, similarly to trained animals, we virally expressed ArchT in D1-Cre or A2a-Cre animals. After 3-5 weeks of viral expression, we implanted a small headpost and a ground pin contralateral to the recorded hemisphere. After allowing animals to recover, a small round craniotomy (1.5mm diameter) was opened over the same coor-

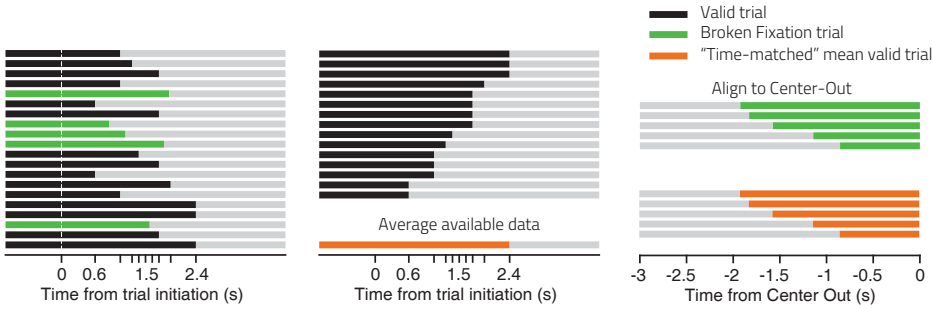


Figure 2.34. Schematic depicting the analysis performed in Figure 2.14 in order to compare activity aligned to broken fixation trials. Briefly, we took valid (black), or broken fixation (green), trials aligned to trial initiation (first tone onset) cropped them at second tone or at the broken fixation event, respectively. We used the valid trials to compute a reference "valid trial" that reflected the average activity of all valid trials, cropped at second tone (mid, orange). Averaging available data (i.e. up until second tone) guarantees that only data from the fixation period is used, without incurring into contamination due to movement onset after the cue is delivered. We subsequently align each broken fixation trial to its occurrence (right) and take, from the reference valid trace, a time matched fragment which we align to the same time since first tone. To compare traces aligned on broken-fixation events to valid trials, we then average all broken-fixation trials and the corresponding time-matched valid reference traces (see Methods).

dinates as the virus injection. Recordings were performed while animals were head-restrained using custom built headbar holders and on top of a passively rotatable cylinder. A silicon probe (ASSY 77-H2, Cambridge NeuroTech) with a tapered optical fibre, identical to the one used for optogenetic inhibition during behavior, glued to the back was slowly lowered into the dorsal striatum. Electrophysiology and laser modulation data were digitized at 30kHz with Open Ephys Acquisition board (Siegle et al., 2017) and acquired with Bonsai(Lopes et al., 2015). Every 10 to 25 seconds, an interval was drawn from the set [0.60,1.05,1.95, 2.40]s and light was continuously delivered during that duration. The range of power used was identical to that used during the task and, similarly, was ramped off for 250ms after the drawn delay had elapsed. In a second batch of animals, we tested the extent of excitation and inhibition when activating medium spiny neurons using ChR2. To achieve ChR2 expression

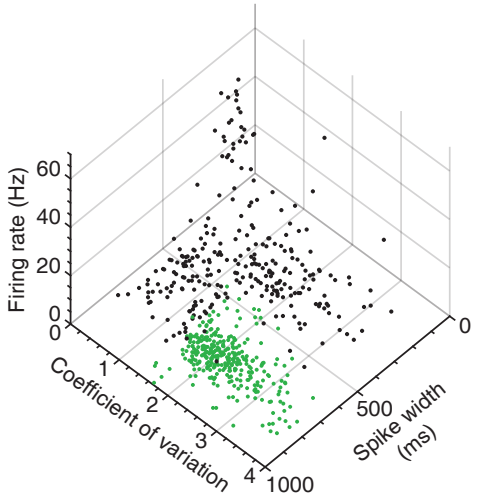


Figure 2.35. Identification of putative medium spiny neurons based on firing statistics and waveform duration (see methods). Green data points indicate units ultimately identified as putative medium spiny neurons (pMSNs).

we virally expressed ChR2 (AAV9-EF1a-double floxed-hChR2(H134R)-EYFP-WPRE-HGHpA). Light intensity was set to 0.5-1mW at the end of the fiber. The remainder of the protocol was identical. Electrophysiology data was sorted using Kilosort2 (github.com/MouseLand/Kilosort2) and manually curated using Phy (github.com/cortex-lab/phy). Spike rates were calculated by binning raw spike counts in 20ms bins. Cells that fired, on average, less than 1/2.4 spikes s^{-1} (detection limit during inhibition) during the baseline period were excluded from analysis. To determine if a cell was modulated, we compared the firing rate in a baseline period prior to light delivery (-3 to -1s) to the average firing rate during the stimuli. We computed a t-test per cell and considered a cell to be significantly modulated if $P \leq 0.05$. To classify units as putative medium spiny neurons we used previously described criteria (Benhamou et al., 2014; Yael et al., 2013) based on firing statistics and spike waveform morphology. We classified units as putative medium spiny neurons if baseline firing rate was less than 10Hz, coefficient of variation greater than 1.4 and waveform duration greater than 500ms (Figure 2.35).

2.4.8 Chronic recordings during task performance & analysis

In order to record from photo identified indirect pathway medium spiny neurons, we crossed A2a-Cre animals with Ai32 (ChR2 reporter line). These animals were trained in identical fashion to the animals used for photometry recordings and optogenetics manipulation. Once stable performance was achieved, we stopped the water deprivation protocol and allowed animals to recover their baseline weight. In the days prior and following surgery we also supplemented their water with Minocycline (10% w/v (Rennaker et al., 2007)). We chronically implanted a single two-shank 64-Channel silicon probe (ASSY-236 H6, Cambridge Neurotech) mounted on a nanodrive shuttle (Cambridge Neurotech). Additionally, we glued a tapered optical fiber (100μm, 0.9mm emission length, 0.22NA, Optogenix) to the back of the probe (200-300μm away from the shanks). The implant was slowly inserted in the dorsolateral striatum, contralateral to a long choice, in the two animals targeting identical coordinates to the ones used for photometry and inhibition experiments. Finally, we chronically cemented a headstage (mini-amp-64 headstage, Cambridge Neurotech) that we connected prior to the start of every session. Animals were given carprofen in the days following surgery and allowed to recover for one week before training resumed. Electrophysiological and synchronization data was digitized at 30kHz with openEphys acquisition board and recorded with Bonsai (Lopes et al., 2015). Data was sorted offline using Kilosort2.5 (github.com/MouseLand/Kilosort) and manually curated using Phy (github.com/cortex-lab/phy). To classify cells as putative medium spiny neurons, we used the same firing statistics and waveform criteria used in the acute recordings. Additionally, we used a photo-identification protocol at the start of each session, consisting of 10 brief pulses, 2ms, of 473nm light at different intensities (2-10 mW measured at the end of the patchcord) and frequencies (5,15 Hz). to photo-identify indirect-pathway medium spiny neurons. Units were considered photo-identified using an intersection of different criteria: SALT test (Kvitsiani et al., 2013) p-value < 0.001, t-test comparing baseline and post laser onset (1-10ms) FR yielding a p-value < 0.01, correlation coefficient between laser-triggered waveform, non-evoked waveform > 0.85 and probability of eliciting $\geq$1

spikes within 1-10ms of each pulse onset > 0.1. For all further analysis, spike counts were binned in 2ms windows and smoothed by convolving with a gamma probability distribution kernel (shape parameter $k = 2$ and scale parameter $\theta =$ 25 ms). To classify neurons on their modulation during the task we compared, for each neuron, the average delay period activity (0.6 to 2.4s) and a corresponding baseline period (-5 to -2s) relative to trial initiation (paired t-test with $p \leq 0.05$ as threshold). Comparison between broken-fixation and valid trials was carried out identically to the photometry data (Figure 2.34). In order to relate the difference in activity between *broken fixations* and valid trials ($\Delta FR_{Broken-Valid}$) and delay period activation, relative to baseline, ($\Delta FR_{Delay-Baseline}$), in Figure 2.15, we took, for each unit, the average $\Delta FR_{Broken-Valid}$ between -0.1 and 0 seconds aligned to broken fixation and regressed it against $\Delta FR_{Delay-Baseline}$. To assess whether this is due to a selection effect dependent on delay period activity, we repeated the same analysis but using a different window aligned to *broken fixations* (-0.5:-0.4 s, slope = -0.042, P = 0.132, t_{87} = -1.522) . Second, we repeated the calculation of $\Delta FR_{Broken-Valid}$ but this time randomly selecting half of the valid trials as a surrogate dataset for *broken fixations* and the remaining half to calculate our valid, null distribution. We bootstrapped the regression slope (against $\Delta FR_{Delay-Baseline}$) from this analysis 1000 times and compared it against our initial outcome (P = 0). Both analyses indicate that our initial result was not due to sporadic correlations in the data or flooring effects from MSNs that were inactive during the delay period.

2.4.9 Optogenetic manipulations during task performance

A 556nm laser (500mW, Optoelectronics Technology) was used as a light source to activate ArchT. Briefly, the output of the laser was aligned to an acousto-optic modulator (AOM MTS110-A3-VIS, AA OPTO-ELECTRONIC) and fibre launched into a patchcord (200 μm core, 0.48NA). The output of the patchcord was then connected to a power splitting rotary joint (FRJ_1x2i_FC-2FC_0.5, DORIC) and finally into 1 or 2 patch cords (200 μm core, 0.48NA) that connected to the animal's implanted tapered fibres (Lambda-B fibre 200μm core, 0.39NA, 1.5mm emitting length, Optogenix). The light intensity was controlled by modulating the AOM using a dedicated Arduino Mega 2560 board connected to a DAC

board (MCP4725, Sparkfun). Regardless of the stimulation duration, all protocols included a 250ms linear ramping off, designed to reduce the potential for rebound excitation (Chuong et al., 2014). The inhibition protocol was applied on a randomly selected 30% of trials. Light was continuously delivered, with the onset aligned to the initiation of a trial (centre nose port) and lasting until second-tone delivery or the exit of the subject from the centre-poke, whichever occurred first. For all analyses of optogenetic manipulation data, the first 15 trials of each session were excluded. Laser power was set to be 23 to 31mW at the end of the fibre for all animals. For a subset of DLS direct pathway animals (n=3) and all DMS animals (n = 6 A2a-Cre and n = 6 D1-Cre), a third protocol was used wherein light was turned on at second tone onset up until choice or 400ms, whichever occurred first. Each day, we changed the location of the inhibition (CS or CL, n=4 DLS A2a-Cre, n=6 DLS D1-Cre, n=6 DMS A2a-Cre, n=6 DMS D1-Cre) and, after collecting data from the unilateral manipulation conditions, we silenced both hemispheres simultaneously (n=4 A2a-Cre and n=4 D1-Cre, n=6 DMS A2a-Cre, n=6 DMS D1-Cre). For single animals, unless otherwise stated, all analyses were performed by concatenating all trials from all sessions of the same manipulation condition.

2.4.10 Movement trajectories

Offline tracking of mouse position was performed using DeepLabCut (Mathis et al., 2018) targeting the nape of each animal (Figure 2.3, Figure 2.24). Low confidence points ($p < 0.5$) were interpolated and tracking data was subsequently smoothed with a 10 sample median filter. To compare movement trajectories between different conditions (Figure 2.24), we first computed the average trajectory of a reference condition (non-inhibited & valid trials). Next, we calculated, for each trial's trajectory, the average euclidean distance to the reference trace. To compute distances between positions from trials of differing duration, the longer trial of the pair was cropped to match the duration of the shorter. To compare across trial pairs of different duration, we normalized euclidean distances by trial duration, computing the average euclidean distance between reference and test trial per sampled frame. We then compared these distributions to the

"null" distribution calculated from the same trials as the average trajectory with a Two-sample Kolmogorov-Smirnov test.

2.4.11 Statistics

Statistical analyses were performed in MATLAB or R. No statistical methods were used to predetermine sample sizes. Unless otherwise stated, data distribution was assumed to be normal. Performed statistical tests, sample size, relevant statistics and p-values are reported throughout the text or in the respective figure caption. We used a linear mixed effects model for the photometry data to account for conditions with missing subject data. We specified random intercepts per animal and tested for fixed effects using the R package lme4(Bates et al., 2015) using data from correct single trials. To test the significance of relevant main effects, we report Anova F-statistics with Satterthwaite adjusted degrees of freedom. We report marginal means and post hoc contrasts (t-statistics), with Tukey correction for multiple comparisons, using the R package emmeans(Searle et al., 1980; Lenth, 2016) as (Effect Size + 95%[CI], p-value) throughout the paper. All tests are two-tailed, unless otherwise stated.

2.4.12 Immunohistochemistry and microscopy

Histological analysis was performed after all experiments to confirm optical fiber placement and expression patterns of transgenes. Mice were administered with a lethal dose of pentobarbital (Eutasil, 100 mg/kg intraperitoneally) and perfused transcardially with 4% paraformaldehyde. The brains were removed from the skull, stored for 24 hours in 4% paraformaldehyde, and then kept in PBS until sectioning. A vibratome or cryostat was used to section the brain into 50 or 40 µm thick slices that were then immunostained with antibodies against GFP (A-6455, 1:1000, Invitrogen) and tdTomato (ab125096, 1:1000, abcam). Finally, all slices were incubated in DAPI. Images were acquired with a confocal microscope (LSM 710, ZEISS) or a slide scanner (Axio Scan.Z1,ZEISS). For electrophysiological recordings, shank placement was confirmed using DiI (V22885, ThermoFisher Scientific).

2.4.13 Reinforcement learning model

We constructed an actor-critic model with the goal of understanding the observed patterns of activity in DLS iMSNs and dMSNs and the functional role of DLS iMSNs in action suppression and selection revealed by optogenetic inhibition experiments. The experimental results demonstrated that the DLS was functionally engaged in suppressing, but not promoting specific actions during the task at times when those actions were most tempting. We thus inferred that some other region must provide the drive to act that constitutes the temptation itself. To accommodate these multiple, apparently regionally distinct, influences on behavior, we designed a model containing two "sub-agents", with access to different state transition dynamics, that interact through behavior in order to solve the task. We start by providing an overview of the sections that describe the model implementation. In the first section: The Interval Discrimination Task, we begin by defining the rules that govern the real world state transitions defined by the behavioral task paradigm itself. However, the model does not operate directly on these real world environmental states, but as described in the section State Representation, contains two internal representations of the external world that are used to learn both state value functions (the critics) and action preference functions (the actors). We then define the general features of the actor-critic architecture in the section titled Actor-critic formulation, describing how the two internal state representations feed into two parallel actor-critic sub-agents, one elaborated with a direct and indirect pathway actor component to model the experimental data involving DLS circuits, and the other modeled more classically without a direct and indirect pathway, for simplicity. In the next section, Learning the values of states, we describe how the sub-agents each learn their own value of expected future reward given the internal state information they have access to using temporal difference learning. Next, in the section Actor learning in a multi-agent setting, we define how a set of action preference functions, that serve as a basis for a policy for behavioral control, are learned in each sub-agent through the influence of a temporal difference RPE that is emitted by each sub-agent's critic. And lastly, we define in the section Policy definition , how these action preference functions emitted by the two sub agents are combined to define a policy for selecting Left, Right, and Hold actions at each time step of a trial.

2.4.13.1 Task environment

The animal's behavior in the task and the underlying neural activity of the striatal populations was modeled as a Reinforcement Learning (RL) problem. RL models require the definition of a Markov Decision Process (MDP) (Sutton et al., 1998) which consists of a set of states $P(e_{t+1}, r_t|e_t, a_t)$ that gives us the probability of moving to a new state e_{t+1} and receiving a reward after performing action .

We model the discrimination task environment by considering a linearly evolving variable $\theta \in \mathbb{R}$ representing the passage of time since the onset of a trial, defined by the onset of the first tone that marked the beginning of interval stimuli. This variable is constrained to the interval $\theta \in [\theta_0, \theta_f]$ where θ_0 is trial initiation/interval onset, and θ_f the time point at which, if no action is taken a timeout is produced and the trial is over. We discretize this interval into a set of N_T bins $e_t \in \mathbb{N}$ of duration $\Delta\theta = \frac{\theta_f - t_0}{N_T}$ in milliseconds. We model the transitions of true time states as an off-diagonal transition matrix $P(e_{t+1}|e_t, a_t)$ were for each true time state $e_t = t$ we move to $e_{t+1} = t + 1$ and where the corresponding true time is given by $\theta = t\Delta\theta$. In each trial, the second tone is defined in a set of M temporal markers $\tau_i \in \mathbb{R}$ all in the range $[\theta_0, \theta_f]$ and given by the values [0.6, 1.05, 1.26, 1.38, 1.62, 1.74, 1.95, 2.4] (s), which will signal the occurrence of the second tone (interval offset) on a given trial i at real time t identified by the true time state e_t.

At each e_t the environment will accept one of 3 actions; choice of the short port ($a = SHORT$), choice of the long port ($a = LONG$) or perform no action, which we define as a HOLD action ($a = HOLD$). If no action other than HOLD is performed during the delay period and the correct choice is made after the second tone, a reward is delivered. If the incorrect action is performed in the post-second tone period or if an action other than HOLD is performed during the delay period (i.e. "Broken fixation"), a punishment is delivered. Reward and punishment are defined as $r_t^+ = 10$ and $r_t^- = -5$, respectively.

2.4.13.2 State representation

Animals including mice are known to express noisy estimates of time(Gibbon, 1977). We model this by considering that the model does not have direct access to the "true time" state e_t but instead relies on an internal estimate of elapsed time given by a set of states for each sub-agent c, $s_t^c \subset [0, ..., N_K] \in \mathbb{N}$ with total number N_K.

This internal estimate was constructed so as to produce scalar variability in temporal estimates around the true time. In order to achieve this, we define a central dwell time μ which corresponds to a sequence of states of each sub-agent s_t^c that accurately follows the elapsed time from the experimentalist's viewpoint, i.e. true time. We then draw the dwell time $\Delta\theta_i^d$, shared by both sub-agents, at each trial i from a gaussian distribution given by $\Delta\theta_i^d \sim \mathcal{N}(\mu, \sigma)$, with variance σ defining the range of potential dwell times. The variable $\Delta\theta_i^d$ thus represents the dwell time for the current trial, becoming discretized in a number of "true time" states, e_t , that elapse in a single internal state, s_t, and is responsible for the scalar variability of internal time estimates expressed by the model across trials. Because the impact of trial to trial dwell time variability accumulates as the model progresses through states, internal variability in time estimates grows in aggregate as true time elapses.

To accommodate assumed differences in information provided by different outlying areas to circuits in distinct regions of the striatum, we constructed different mappings from environmental to internal states, K^c for each sub-agent $c \in \{D/I, \chi\}$ (Equation 2.3, Equation 2.4). Each sub-agent maps the true time state e_t to the internal temporal estimate of the agent given by the state s_t^c. The function K^c maps the internal temporal estimate into two possible sets of states; post 1st tone states (or pre 2nd tone states) $S^{pre} \in [0, ..., N_K/2] \subset \mathbb{N}$ and post 2nd tone states $S^{post} \in [N_K/2, ..., N_K] \subset \mathbb{N}$ with the transition between these two sets triggered by the occurrence of a tone at true time τ_i. Sub-agent $c = D/I$ is intended to reflect circuitry that includes dMSNs and iMSNs in DLS, whereas sub-agent X is intended to reflect circuitry elsewhere. This distinction between

the sets $\{S^{pre}, S^{post}\}$ is what allows the model to correctly discern in which states it should perform an action or suppress all actions through the HOLD action.

The mapping $K^{D/I}$, corresponding to sub-agent $\{D/I\}$, produces a more veridical representation of the task environment, including whether a second tone has occurred and thus is defined in the the full set of states $K^{D/I} : \{\theta, \Delta\theta_i^d, \tau_i\} \rightarrow \{S^{pre}, S^{post}\}$. Trial start commences a progression through post-first tone states (i.e., pre-second tones states), and occurrence of the second tone causes a transition into a stream of post-second tone states, through which the state progresses until a terminal state is reached when a choice is made or the maximum possible state in the trial is reached.

We define the mapping concretely as the sequence of states given by

$$K^{D/I}(\theta, \Delta\theta_i^d, \tau_i) = \begin{cases} \lfloor \frac{\theta}{\Delta\theta_i^d} \rfloor & \text{if } \theta < \tau_i \\ \\ \lfloor \frac{\theta}{\Delta\theta_i^d} \rfloor + J & \text{if } \theta \geq \tau_i \end{cases} \tag{2.3}$$

with $\lfloor x \rfloor$ the floor function applied to x and J a jump parameter that goes from S^{pre} to S^{post} states. The second mapping K^X, corresponding to sub-agent X, is unaware of second tone events and is thus only defined in pre-second tone states, *i.e.* $K^{\chi} : \{\theta, \Delta\theta_i^d\} \rightarrow S^{pre}$. In this mapping, as in the previously described one, trial start commences a progression through post-first tone states (i.e., pre-second tone states), however, unlike the previous mapping, progression through pre-second tone states continues until a terminal state is reached either when a choice is made or when the maximum possible state in the trial is reached, regardless of second tone occurrence. We view the mapping K^X as more abstract, useful for separating the world into short ($<$1.5s) or long ($>$1.5s) epochs and perhaps constituting a more predictive state representation(Dayan, 1993a; Stachenfeld et al., 2017), yet devoid of more low-level sensed features of the environment. Mapping K^X also reflects a kind of information factorization, hiding information about the environment from sub-agent X in a manner hypothesized

within hierarchical views of behavioral control to aid in generalizability(Merel et al., 2019).

Concretely, we define this second mapping for each trial through the function

$$K^{\chi}(\theta, \Delta\theta_i^d) = \lfloor \frac{\theta}{\Delta\theta_i^d} \rfloor \tag{2.4}$$

The full set of mappings for the set of sub-agents is thus given by:

$$s_t^{D/I} = K^{D/I}(\theta, \Delta\theta_i^d, \tau_i) \tag{2.5}$$

$$s_t^{\chi} = K^{\chi}(\theta, \Delta\theta_i^d) \tag{2.6}$$

2.4.13.3 Actor-critic formulation

To construct each sub-agent we formulate an actor-critic algorithm where, for each state representation $\{s_t^{D/I}, s_t^{\chi}\}$, we define a corresponding critic $V^c(s_t^c)$ with index $c = \{D/I, \chi\}$ and its respective actor(s)(Motiwala et al., 2020) defined by a set of action preferences(Sutton et al., 1998) functions $A^p(s_t^c, a_t)$ where the index $p = \{D, I, \chi\}$ corresponds to the direct, indirect pathways and the non-DLS actor, respectively. Sub-agent D/I is constructed to contain direct and indirect pathway actors with opposing influence on action production, whereas sub-agent X is constructed with a more classical, single actor, for simplicity. The learned Action Preferences are then combined into a single probabilistic policy $\pi(a|s_t^{D/I}, s_t^{\chi})$ from which a decision will be sampled.

The choice of action preference algorithms for this problem is twofold; actor-critic models have been mapped onto striatal anatomy, where striatal actors are taught by a critic that emits reward prediction errors through dense dopaminergic projections to the striatum(Bornstein & Daw, 2011). Secondly, due to the fact that the action preference functions are learned using the reward prediction error δ_t, we can define update rules onto direct and indirect pathway actors based on previously established effects of dopamine on plasticity of synapses on striatal

dMSNs and iMSNs (Collins & Frank, 2014; K. N. Gurney et al., 2015; Iino et al., 2020; S. J. Lee et al., 2021).

2.4.13.4 Learning the value of states

In the usual actor-critic formulation, the critic takes the form of a state value function, defined as the expected value over a policy $\pi(a|s_t)$ of the sum of discounted rewards and is written as ,

$$V(s_t) = \mathrm{E}^{\pi}\left[\sum_{k=0}^{T-k-1} \gamma^k r_{t+k+1} \middle| s_t\right] \tag{2.7}$$

with $\gamma < 1$ being the discount factor.

In order to update the state-values, Temporal Difference (TD) methods are used (Sutton et al., 1998), with an update given by:

$$V(s_t) \leftarrow V(s_t) + \alpha\delta_t. \tag{2.8}$$

where α is the learning rate for the critic, V for value function, and the *Reward Prediction Error* δ_t (RPE) defined as:

$$V(s_t) \leftarrow V(s_t) + \alpha\delta_t. \tag{2.9}$$

In our formulation, the RPE will serve as a learning signal for all the actors that will be defined below. In order to make the notation explicit for the multi-agent formalisation that we'll expand in the following sections, we define that for each critic $V^c(s_t^c)$ their corresponding value function is updated by a corresponding RPE δ_t^c so the update rules for an agent consisting of multiple critics can be written as,

$$V^c(s_t^c) \leftarrow V^c(s_t^c) + \alpha(s_t^c)\delta_t^c \tag{2.10}$$

$$\delta_t^c = R_{t+1} + \gamma V^c(s_{t+1}^c) - V^c(s_t^c) \tag{2.11}$$

with γ the discount factor of the sub-agents and the learning rate is defined as the inverse of the number of visits to state s_c^t, $N(s_t^c)$ (Sutton et al., 1998; Doya, 1999), *i.e.*:

$$\alpha(s_t^c) = \frac{1}{N(s_t^c)} \tag{2.12}$$

in order to facilitate convergence.

2.4.13.5 Actor learning in a multi-agent setting

In order to model the neural activities of the dMSNs and iMSNs observed in the data we take a Multi-Agent RL approach(Buşoniu et al., 2010) by hypothesizing that the model consists of a set of critics feeding information to different actors that represent the underlying neural activity of each striatal pathway. For this formalism, each actor can have a different viewpoint as to what is happening both in terms of their provided internal state representation and in the corresponding RPEs that each receives, being sensitive to a specific range of the available RPEs, as was previously done in Collins & Frank (2014); K. N. Gurney et al. (2015).

We take inspiration from biology and the known results regarding the underlying plasticity dynamics in striatal neurons so as to model the activities of the direct and indirect pathways. Based on a number of experimental studies, K. N. Gurney et al. (2015) posit that positive changes in overall plasticity of direct pathway medium spiny neurons (dMSNs) are directly correlated with an increase in dopamine (DA) whilst decreases in DA concentration show a decrease in plasticity. Conversely, for indirect pathway MSNs (iMSNs) higher DA values have the opposite effect as they reduce the change in plasticity while lower concentrations of DA potentiate plasticity. Consistent findings have also been observed in the volume change of dendritic spines for dMSNs and iMSNs (Iino et al., 2020), spike timing dependent plasticity(Shen et al., 2008), well as intracellular PKA levels (S. J. Lee et al., 2021). In addition, these two populations appear largely spatially intermixed and thus likely to receive comparable dopamine input. We capture these plasticity rules by defining a set of nonlinear functions f^p

that we'll associate with individual actors p. As the amount of DA present in the MSNs modulates the change in plasticity of these cells, we'll make this nonlinear function dependent on the RPE of its corresponding sub-agent c,*i.e.* $f^p(\delta_t^c)$. In this manner we're able to associate a given pathway to a specific combination of a nonlinear function f^p and the RPE of a given critic δ_t^c.

Concretely, the direct pathway actor of the DLS sub-agent is subject to updates via $f^D(\delta_t^{D/I})$ and the indirect pathway actor of the DLS sub-agent via $f^I(\delta_t^{D/I})$. These functions are defined for each pathway p as,

$$f^p(\delta_t^c) = b_0^p + \frac{b_1^p}{1 + b_2^p e^{(1-b_3^p \delta_t^c)}} \tag{2.13}$$

where the b parameters are adjusted for each individual pathway. Concretely, b_0^p and b_1^p determine the vertical range of the curve and b_2^p, b_3^p it's slope. For the direct, indirect and X pathways we used [-0.7, 8, 3.7, 1], [-0.7, 8, 3.7, -1] and [-4, 8.5, 0.45,1] respectively. The parameters of the indirect pathway are defined so as to transform the negative RPEs into positive update weight changes, the direct pathway function transform positive RPEs into positive the corresponding weight changes for that actor and the second sub-agent transfer function is defined so that it can assume the joint functionality of the two transfer functions that exist in the D/I agent, transforming both positive and negative RPEs into their corresponding positive and negative update weights with a single function.

In actor-critic formalism one needs to learn the policy in addition to the state value function. In our case we chose the action preference framework, with action preferences defined as functions $A(s_t, a_t)$ at each state-action pair that will allow the agent to choose the appropriate action at each state. In the usual definition of actor-critic models, for an action a_t and state s_t the action preferences are learned through the RPE that is calculated at the current time, δ_t.

The update rule for the action-preference for a single actor receiving RPEs from a critic is given by

$$A(s_t, a_t) \leftarrow A(s_t, a_t) + \alpha\delta_t \tag{2.14}$$

In simple terms, at each update if the action executed at state s_t results in a positive RPE then that action will become more likely to be performed in the future, if the RPE is negative then the likelihood of the action decreases for that state.

Inserting the aforementioned plasticity rules to drive action-preference change in the model is done by slightly altering the action-preference updating learning rules defined above. We apply the pathway transfer function $f^p(\delta_t^c)$ (defined in Equation 2.13) to each actor's learning rule, i.e.

$$A^p(s_t^c, a_t) \leftarrow A^p(s_t^c, a_t) + \alpha(s_t^c) f^p(\delta_t^c) \tag{2.15}$$

and define the learning rate for the actors in a similar manner to the critics.

Concretely, for the actors and state representations used in our model, the full set of update equations is thus:

$$A^D(s_t^{D/I}, a_t) \leftarrow A^D(s_t^{D/I}, a_t) + \alpha(s_t^{D/I}) f^D(\delta_t^{D/I}) \tag{2.16}$$

$$A^I(s_t^{D/I}, a_t) \leftarrow A^I(s_t^{D/I}, a_t) + \alpha(s_t^{D/I}) f^I(\delta_t^{D/I}) \tag{2.17}$$

$$A^\chi(s_t^\chi, a_t) \leftarrow A^\chi(s_t^\chi, a_t) + \alpha(s_t^\chi) f^\chi(\delta_t^\chi) \tag{2.18}$$

with the RPEs for each agent being given by two independent critics with update rules:

$$V^{D/I}(s_t^{D/I}) \leftarrow V^{D/I}(s_t^{D/I}) + \alpha(s_t^{D/I})\delta_t^{D/I} \tag{2.19}$$

$$V^\chi(s_t^\chi) \leftarrow V^\chi(s_t^\chi) + \alpha(s_t^\chi)\delta_t^\chi \tag{2.20}$$

2.4.13.6 Policy definition

In order to calculate the policy as a combination of the 3 pathways we define a total action-preference for the three actors as a sum of their corresponding action preferences

$$A^T(s_t^{D/I}, s_t^{\chi}, a_t) = \omega_D A^D(s_t^{D/I}, a_t) + \omega_I A^I(s_t^{D/I}, a_t) + \omega_\chi A^\chi(s_t^{\chi}, a_t) \quad (2.21)$$

where the weights are given by $\omega_D = -\omega_I = \omega_\chi = 1$. Because the indirect pathway actor has an opposite sign to the remaining sub-agents, it serves as an inhibitory gate for action execution. This idea was first proposed in Mink (1996), where it is highlighted that the expression of movement seems to be directly correlated with the amount of downstream inhibition exerted by outputs of the basal ganglia. Later experiments using optogenetics then showed that increased activity in D2-MSNs (iMSNs) via optogenetic excitation generated a decrease of movement in mice, whilst an increase of activity in D1-MSNs (dMSNs) generated an increase in movement (Kravitz et al., 2010).

All of the suppressive information is taught by negative RPEs acting on the A^I action-preferences. One way of looking at this function is that it represents, instead of the policy which tells agents what to do, an anti-policy that informs the model of what not to do. It serves as a dynamic map that evolves through the environment, releasing and increasing inhibitory pressure on the available positively tuned signals that push for actions to be expressed.

It is through the value of this total action preference that we define the HOLD action. Given that suppressing actions in animals is linked with higher values of indirect pathway activity, we increase the probability of that action being performed proportionally to how negative the total action preference is, as negative total action preferences are mostly a result of a dominance of the indirect pathway actor over the remaining actors. We achieve this by, instead of learning an explicit action preference for the HOLD action as we do for the SHORT or LONG actions, triggering a condition where if the value of the total action preference is negative, the action-preference of the HOLD action will be zero. Concretely,

$$
\begin{aligned}
if \quad A^T(s_t^{D/I}, s_t^{\chi}, a_t) < 0 \quad \forall \quad a_t = \{SHORT, LONG\} \\
\Rightarrow A^T(s_t^{D/I}, s_t^{\chi}, HOLD) = 0 \\
else \\
\Rightarrow A^T(s_t^{D/I}, s_t^{\chi}, HOLD) = -\infty
\end{aligned}
\tag{2.22}
$$

This definition for the HOLD action action preference works in the following way: In the usual actor-critic formulation, given the discrete states s_t, the probability of the model choosing one of the actions a at a given state is given by the distribution $\pi(a|s_t)$. For a single actor, the actions at each observation s_t are chosen via a softmax of the action-preference $A(s_t, a_t)$, *i.e.* the probability of a given action a being executed when an agent observes a state s_t can be obtained by sampling from the distribution given by

$$
\pi(a|s_t) = \frac{e^{\beta A(s_t,a)}}{\sum_j e^{\beta A(s_t,a^j)}} \tag{2.23}
$$

where β is the temperature parameter controlling exploration, and the denominator contains the sum over all actions.

In our multi-agent model, we define the probability of choosing action a given states $\{s_t^{D/I}, s_t^{\chi}\}$ for the actions as the softmax of the total action preferences $A^T(s_t^{D/I}, s_t^{\chi}, a_t)$

$$
\pi(a|s_t^{D/I}, s_t^{\chi}) = \frac{e^{\beta A^T(s_t^{D/I},s_t^{\chi},a)}}{\sum_j e^{\beta A^T(s_t^{D/I},s_t^{\chi},a^j)}} \tag{2.24}
$$

with the conditions for the HOLD action defined in Equation 2.22. With these conditions, in the case where the total action preference for both SHORT and LONG actions is negative and suppression dominates, the term in the denominator for the HOLD action will be one and terms for the remaining actions will have small values as the exponential function approaches zero for negative values. For the case where suppression does not dominate, only the SHORT and

LONG actions have probabilities bigger than zero, as the action-preference for the HOLD action goes to $-\infty$ (and probability zero).

2.4.13.7 Simulation details

All simulations were performed on an 80 dimensional state-space. Each trial consisted on a trajectory with a pre-specified dwell time sampled from a gaussian distribution with mean $\mu_t = 67$ ms and variance $\sigma_t = 20$ ms, sampled from a predetermined set of 30 dwell times that covered the interval between $t_max =$ 100 ms and 65 ms. Convergence was assessed by verifying that all value functions and action preferences showed a normalized difference between epochs under 0.01 for more than 50,000 trials, with a total number of 250 000 trials. The training set for the second tone times (in seconds) was given by $T_{train} =$ [0.1, 0.3, 0.5, 0.6, 0.8, 1.05, 1.26, 1.38, 1.62, 1.74, 1.95, 2.2, 2.4, 2.5] with the decision boundary being set at 1.5s. The learning rates were all initialized at 20 and decreased for each state proportionally to the number of visits to each state. Rewards for correct and incorrect/broken fixation trials were set at 10 and -5, respectively.

The transfer function $f^p(\delta_t^c)$ parameters were defined as $[b_0^p, b_1^p, b_2^p, b_3^p]$. The values for the direct pathway actor were set at [-0.7, 8, 3.7, 1], [-0.7, 3.7, 8, -1] for the indirect and [-4, 8.5, 0.45,1] for the third actor. The weights for each actor $\omega_D = -\omega_I = \omega_\chi$ were all equal to 1. The discount parameter for each critic γ was 0.98 and the temperature of the softmax function β for the policy was 1.5 for both training and testing.

In order to simulate the effects of pathway specific optogenetic inhibition, we generated control and manipulated datasets (1.5k trials) drawing single trial intervals from a set identical to the one used to test the animals (T_{test} [0.6, 1.05, 1.26, 1.38, 1.62, 1.74, 1.95, 2.4]).

2.4.13.8 Optogenetics experiments

For all of the post-training experiments the learning of the agents was frozen and for the control trials the weights of each agent were kept the same as during training. Two sets of experiments were performed; one where the indirect-pathway

was inhibited and another where the third agent X was inhibited. This inhibition consisted in a downscaling of the weights of the corresponding pathway weights to $\omega_I = 0.93$ and $\omega_X = 0.98$ so as to match the effect size observed in the data. Varying these values produces qualitatively similar effects. For each perturbed pathway, three optogenetic perturbations were performed, corresponding to a unilateral or bilateral manipulation where the weights corresponding to the relevant action (SHORT, LONG or SHORT and LONG) were scaled to the values defined above.

2.4.13.9 Pseudo-code

Algorithm 1 Opponent Multi-Agent Actor-Critic Algorithm

Define $\{\beta, \gamma, \mu_t, \sigma_t, b_i^p, \omega^p, \tau_i\}$
Initialize states $\{s_t^{D/I}, s_t^{\chi}\}$
Zero initialization of critics $\{V^{D/I}(s^{D/I}), V^{\chi}(s^{\chi})\}$
Zero initialization of action preferences
$\{A^D(s^{D/I}, a), A^I(s^{D/I}, a), A^{\chi}(s^{\chi}, a), A^T(s^{D/I}, s^{\chi}, a)\}$
Zero initialization of state visits $N(s_t^c)$
for *episode* **do**
    Sample dwell time $t_d \sim \mathcal{N}(\mu_t, \sigma_t)$
    for *step* in *episode* **do**
        Update Learning rates:
        $\alpha(s_t^{D/I}) \leftarrow \alpha(s_t^{D/I}) + (N(s_t^{D/I}) + 1)^{-1}$
        $\alpha(s_t^{\chi}) \leftarrow \alpha(s_t^{\chi}) + (N(s_t^{\chi}) + 1)^{-1}$
        Choose action and get next state:
        Check if HOLD action should be triggered:
        if $A^T(s_t^{D/I}, s_t^{\chi}, a) < 0 \quad \forall \quad a = \{SHORT, LONG\}$ **then**
            $A^T(s_t^{D/I}, s_t^{\chi} HOLD) = 0$
        else
            $A^T(s_t^{D/I}, s_t^{\chi}, HOLD) = -\infty$
        end if
        Sample new action $a_t \sim \pi(a|s_t^{D/I}, s_t^{\chi})$
        Observe $\{s_{t+1}^{D/I}, s_{t+1}^{\chi}, r_t\}$ from $\{K^{D/I}(t, \tau_i, t_d), K^{\chi}(t, t_d)\}$
        Update Critic:
        $\delta_t^{D/I} = r_t + \gamma V(s_{t+1}^{D/I}) - V(s_t^{D/I})$
        $\delta_t^{\chi} = r_t + \gamma V(s_{t+1}^{\chi}) - V(s_t^{\chi})$
        $V(s_t^{D/I}) \leftarrow V(s_t^{D/I}) + \alpha(s_t^{D/I})\delta_t^{D/I}$
        $V(s_t^{\chi}) \leftarrow V(s_t^{\chi}) + \alpha(s_t^{\chi})\delta_t^{\chi}$
        Update Action Preferences:
        $A^p(s_t^D, a_t) \leftarrow A^p(s_t^D, a_t) + \alpha(s_t^{D/I}) f^D(\delta_t^{D/I})$
        $A^p(s_t^I, a_t) \leftarrow A^p(s_t^I, a_t) + \alpha(s_t^{D/I}) f^I(\delta_t^{D/I})$
        $A^p(s_t^{\chi}, a_t) \leftarrow A^p(s_t^{\chi}, a_t) + \alpha(s_t^{\chi}) f^{\chi}(\delta_t^{\chi})$
        $A^T(s_t^{D/I}, s_t^{\chi}, a_t) = \omega_D A^D(s_t^{D/I}, a_t) + \omega_I A^I(s_t^{D/I}, a_t) + \omega_{\chi} A^{\chi}(s_t^{\chi}, a_t)$
    end for
end for

Chapter 3

General discussion

3.1 Main findings

One basic problem brains must solve is how to achieve the best outcome in the face of an ever-changing world. To achieve this goal, brains must rely on a stream of incoming sensory information, internal state, and past experience in order to learn and adapt. However, exactly how this organ achieves such a marvelous feat is far from well understood.

The basal ganglia, in particular, are involved in several functions thought to be critical for a myriad of forms of adaptive behavior. They have been implicated in regulating low-level motor output, sensory processing, motivation, attention, learning, memory, decision making, and ultimately, the overall process of action selection (Doya, 1999; K. Gurney et al., 2001). In the work presented in this dissertation I studied how distinct BG circuits support different aspects of action selection.

First, we provide compelling evidence for the functional opposition between the two major feedforward circuits in BG: the direct and indirect pathways. Both anatomical and physiological data support this classical conceptual model (DeLong, 1990), however previous reports of large-scale co-activation during movement have put into question this dichotomous view (Cui et al., 2013; Tecuapetla et al., 2014; Cox & Witten, 2019). Critically, while we do reproduce the finding of general co-activation during movement, we also show that by training animals to adopt a state of prolonged action suppression we can observe signs of opponency in the endogenous activity of the two pathways in sensorimotor striatal circuits (DLS). Our findings are consistent with the reconciling theory that action production progresses through the simultaneous recruitment of action promotion and suppression processes (Mink, 1996).

Second, we show that, in a task where animals must alternate between periods of action production and suppression, DLS is largely engaged to support the latter. In particular, indirect pathway activity in a given hemisphere qualitatively follows the need to suppress a contra-lateral action during those times when such a movement is potentially rewarded and thus tempting. Moreover, by optogenet-

ically silencing this population of neurons we found that such engagement not only reflects this need, but it is also necessary to determine *whether*, *when* and, *which* actions animals are able to suppress.

Lastly, we reasoned that the dynamics associated with action suppression in DLS arise to counteract a respective action-promoting drive carried out by circuits *elsewhere* in the brain. With this intuition in mind, we built a simplified, yet biologically inspired, multi-agent reinforcement learning model capable of reproducing our main behavioral, physiological and optogenetic manipulation results. Moreover, following a prediction generated from the model, we found that dorsomedial striatal (DMS) circuits are part of the "*elsewhere*" circuit responsible for carrying the action-promoting drive that ultimately leads to the premature behavior observed in our task.

This final chapter will be dedicated to the discussion of how our findings relate to the current view of BG function, and how they might inform new conceptual models and experiments. As is often the case, our work will originate more questions than those it answers. Hence, I feel I should warn the reader that despite the fact that I will try to support my arguments with data from our own experiments and background presented in Chapter 1, the discussion will also be seasoned with a healthy amount of speculation and currently untested hypotheses.

3.2 On the (seeming) suboptimality of behavior

When studying animal behavior we often assume that organisms evolved to behave optimally with respect to some goal, for instance: finding food, shelter and mating partners. This axiom is a staple of many foundational learning theories (Thorndike, 1911; Skinner, 1938). However, in several cases, it is possible to observe animal behavior that is seemingly incompatible with such an assumption. In one of their papers, Keller and Marian Breland, anecdotally describe their observations when attempting to train a wide variety of animal species in many different operant conditioning paradigms:

"When we began this work, it was our aim to see if the science would work beyond the laboratory, to determine if animal psychology could stand on its own feet as an engineering discipline. These aims have been realized.[...] Emboldened by this consistent reinforcement, we have ventured further and further from the security of the Skinner box. However, in this cavalier extrapolation, we have run afoul of a persistent pattern of discomforting failures. These failures, although disconcertingly frequent and seemingly diverse, fall into a very interesting pattern. They all represent breakdowns of conditioned operant behavior.[...] The examples listed we feel represent a clear and utter failure of conditioning theory. They are far from what one would normally expect on the basis of the theory alone. Furthermore, they are definite, observable; the diagnosis of theory failure does not depend on subtle statistical interpretations or on semantic legerdemain—the animal simply does not do what he has been conditioned to do."

Breland & Breland (1961)

In many of the examples provided in their manuscript, animals are shown to be aware of the instrumental contingencies of the task and yet, despite long painstaking months of training, keep resorting to maladaptive behavior. Once one acknowledges the existence of such a seemingly irrational behavior - *i.e.,* that which follows a suboptimal decision making process - it immediately raises the question of what are the constraints and/or computations that give rise to these behavioral policies.

In Chapter 2 of this dissertation I introduced a novel behavior paradigm that showcases such an example of, in light of the imposed task structure, suboptimal behavior in a significant proportion of the trials. In the presented paradigm, under no circumstance will prematurely leaving the centre-nose port increase the total amount of rewards collected. Moreover, despite being immediately informed that such a premature action will result in no reward for a given trial (through the delivery of a salient white-noise burst auditory cue, that is present from the first session of training), animals nevertheless try to report their choice at one of the two side ports in the majority of trials (Figure 2.4). How to explain such deviation from optimal behavior?

First, I would argue that *optimality* as defined by the maximization of an expected utility leads to a correspondingly narrow definition of *optimal* behavior (Charnov, 1976). While artificial agents might indeed be constructed to behave in an optimal manner when performing a specific task and under a precisely defined set of rules and goals, biological agents are not afforded such an intelligent design. Instead animal behavior is also, to some extent, a function of evolutionary processes that gave rise to all the *hardware* available for computing and interacting with the world (Zador, 2019). Therefore, when one tries to construct, even qualitatively, the objective function that animals try to maximize, it is important to not only consider specific costs and benefits but also the genetic "baggage" accumulated through the evolutionary process (Cisek, 2019; Bergman & Beehner, 2021). This is all the more critical when one considers that evolution likely traverses a tortuous fitness landscape that might result in strategies, or heuristics, that are often "just good enough" (J. M. Smith, 1979; Brooks, 1991).

As a toy example, consider a set of hypothetical biological agents that, across several generations, occupy a particular *niche* where the color red is often associated with tasty, nutritious berries. During the evolutionary process, visual circuits that are able to bias approaching behavior towards small round red objects might have been systematically selected, giving rise to what an external observer would likely consider an optimal behavior (*i.e.,* approach red berries). However, should one take this same animal and displace it to a distinct environment wherein visually identical berries would now be poisonous, such a heuristic would undoubtedly give rise to what the same observer would consider deleterious behavior. In other words, given a specific *niche* that biological agents occupy for several generations, some degree of ethnologically relevant neuronal and behavior specialization is likely to emerge in the population (von Uexküll, 2013; Davies et al., 2012). We, as observers, should thus tread carefully when interpreting animal behavior in light of some presumed local optimality constraint.

In the specific case of our paradigm, it is possible that premature behavior displayed by mice is the result of a similarly selected useful heuristic (Marsh, 2002). What evolutionary pressure might have biased the selection of *prematureness* as a useful strategy? One can only speculate, but is possible that in environments

wherein 1) competition between individuals is rampant, 2) the safety cost of waiting, or standing still, for large periods of time is high (*e.g.,* due to the presence of predators) or 3) the statistics of the environment change quickly relative to the time-scale of a decision making process, *prematureness* might turn out to be a useful heuristic to improve the likelihood of survival and eventually passing on genetic material to future generations.

Relatedly, *hardware* constraints might also shape behavior in a different way. Cognitive resources such as attention, working memory, and processing power are limited, and thus likely to be exhausted under particularly demanding tasks (Broadbent, 1965). Given how long it takes to train animals in our paradigm, it might be possible that we are approaching a specific resource limit available to this species, that eventually give rise to this form of suboptimal behavior. Anecdotally, rats trained in an identical version of the task in our lab (Monteiro et al., 2020), display dramatically lower rates of premature behavior when compared to mice.

Meanwhile, brains, and their functionally specialized sub-systems, likely evolved in a context where animals must not only satisfy a single goal but instead be robust to a variety of problems they might face in the wild. Such an ability to generalize is expected to come at the cost of local optimality (Davies et al., 2012).

Finally, as we show at the end of Chapter 2, the behavior output in our task can be seen as the result of the interaction between multiple agents operating in parallel to generate behavior. Once again, while it might initially appear counter-intuitive to possess an arrangement where distinct agents, with potentially different views on the world - and thus pursuing distinct drives to act -, it has been shown to be advantageous in a wide variety of circumstances (Merel et al., 2019). An important open questions still remains: what makes the potential interaction between sub-agents stochastic (*i.e.,* animals only *break fixation* in some trials)? Similarly to a probabilistic process of decision making, it is possible that small fluctuations in the dynamic process of either of the agents, or simply neuronal

noise in the overall process, results in sporadic premature behavior (Busemeyer & Townsend, 1993; Binder et al., 2004; Mountcastle et al., 1975).

At the end of the day, cases of irrational behavior, with clear deviations from optimality, present a unique opportunity to study the mechanistic constraints of neuronal computations, and how these ultimately lead to behavior.

3.3 Direct and indirect pathways, revisited

3.3.1 On the functional opponency of the direct and indirect pathways

Initial anatomical and physiological observations led to the hypothesis that the two major pathways of the basal ganglia play functionally opposing roles. Further experiments where the activity of these two pathways was selectively manipulated did indeed confirm that direct and indirect pathways are capable of exerting broadly opponent influence on various aspects of behavior such as locomotion (Kravitz et al., 2010; Roseberry et al., 2016), ongoing motor sequence production (Tecuapetla et al., 2016; Sippy et al., 2015), reinforcement (Kravitz et al., 2012; Yttri & Dudman, 2016) and decision making (Tai et al., 2012). However, in apparent contradiction with this classic view (Tecuapetla et al., 2014, 2016; Cox & Witten, 2019), wherein these two pathways differentially regulate behavior by opposingly modulating the BG output nuclei, concurrent activation is often generally observed during movement (Cui et al., 2013; Tecuapetla et al., 2014; Markowitz et al., 2018). A long-standing, and potentially reconciling, hypothesis proposes that during normal behavior action proceeds by promoting a given motor plan while simultaneously suppressing other, potentially competing, ones (Mink & Thach, 1993).

This concept seeded the design of the paradigm presented in Chapter 2. We wondered under which behavior context would one be able to observe signs of functional opponency in the endogenous activity of the two pathways. We reasoned that it should not be while animals are actively performing action but instead in a situation where behavior is largely governed by a state of action

suppression and active immobility. Under such a scenario we hypothesized that the activity of the indirect pathway should be relatively higher than that of the direct pathway.

We thus recorded from the dorsolateral striatum, an area where co-activation was previously observed (Cui et al., 2013; Tecuapetla et al., 2014; Markowitz et al., 2018). Consistent with previous studies, we were able to observe the co-activation of the two pathways during movement, by looking at epochs during which animals are required to perform actions, specifically when initiating a trial and reporting their choice at one of the side ports, respectively (Figure 2.9, Figure 2.33). Critically, during the period of active immobility, activity in the indirect pathway was sustained whereas direct pathway acivity levels were relatively lower. Moreover, by optogenetically silencing DLS iMSNs, we showed that this preferential engagement was not only a reflection of the putative need to suppress action but was instead causally involved in supporting it (Figure 2.18). This is, to the best of our knowledge, the first time large-scale signs of functional opponency have been observed in the endogenous pattern of activity in the two classes of medium spiny neurons.

Interestingly, while we found signs of functional opponency between the two pathways in DLS, some degree of functional asymmetry on the effects of optogenetic inhibition was concurrently observed. Specifically, silencing DLS iMSN activity reduced animals' ability to suppress a contra-lateral action during the period of active immobility but did not change the vigor with which these choices were performed. Conversely, silencing DLS dMSN activity did not alter the probability of premature behavior but instead increased the time animals took to report a choice after *breaking fixation*, consistent with previous reports implicating BG in regulating action vigor (Turner & Desmurget, 2010; Panigrahi et al., 2015; Dudman & Krakauer, 2016).

It is possible that, due to technical limitations, such as flooring/ceiling effects related to the already low activity observed in direct pathway photometry recordings during the delay period, potential positive results of the optogenetic manipulation failed to be observed. However, similar reports of asymmetric effects on

behavior have been also reported elsewhere when manipulating two populations of MSNs (Tecuapetla et al., 2016; Geddes et al., 2018; Peak et al., 2020). For instance, Tecuapetla et al. (2016) report that direct pathway silencing slowed the initiation of an action sequence, while manipulating the iMSN population led to abortion of the already ongoing action sequence. Interestingly, the latter result is also seemingly compatible with our interpretation of indirect pathway function, wherein lowering the suppression of other actions might lead animals to *switch* (*i.e.,* engage in other behaviors) prematurely.

In general, care must be taken when interpreting the effect of manipulating these two pathways during behavior. In Tecuapetla et al. (2016) authors acknowledge that, at first glance, manipulating direct or indirect pathway activity leads to an increased latency to initiate a trial. However, as one might expect, there are several ways in which behavior can be affected resulting in an increase of latency to initiate a trial (*e.g.,* decrease in overall vigor, uncoordinated motor output, *dithering*, etc.). These potential confounds highlight the need for carefully designed behavioral paradigms that are able to probe the contribution of neuronal substrates to specific computations and behaviors. In our case, the observed vigor decrease is present in the apparent absence of choice effects and vice-versa.

Additionally, it should be noted that the effects of manipulating these two populations will likely vary depending on the unique demands of the task, or behavior, being studied. Therefore, some caution is advised when generalizing the outcome of these manipulations to other behaviors. For example, while in our experiments we failed to detect changes in choice behavior when inhibiting the direct pathway of the DLS, equivalent manipulations of DMS circuitry resulted in ipsilateral bias following *broken fixations'* choices. This finding highlights that despite operating in general functional opposition, spatially localized circuits can be selectively recruited to generate adaptive behavior. Similarly, a recent study by Bolkan et al. (2021) found that medial circuits of the BG are necessary during a task that was designed to leverage mice working memory, but not during an identical version that relies on overt sensory cues to collect rewards. This interpretation is consistent with both our model and data in that more medial

circuits are thought to receive higher-order inputs from frontal cortical areas that are thought to support working memory (Postle, 2006) and thus able to provide such information to downstream striatal circuits that can be used to bias the decision-making process.

Finally, the prevalence of reports of symmetrical effects when optogenetically activating either one of the two populations (*e.g.,* (Kravitz et al., 2010, 2012; Tai et al., 2012; Sippy et al., 2015; Yttri & Dudman, 2016)) but not when silencing their activity ((Tecuapetla et al., 2014, 2016; Delevich et al., 2020)), raises the interesting point that while increasing the activity of either pathway is *sufficient* to alter different parameters of behavior, it might not inform as to whether the circuit is engaged under normal conditions, and whether they are thus *necessary* to support action. Indeed, the authors in Bolkan et al. (2021) make a similar point: *"These negative findings argue against a major involvement of endogenous activity in DMS pathways in the execution of movement in the absence of a decision. This is consistent with the dearth of reports demonstrating strong and opposing modulation of behavior by striatal pathways using pathway-specific optogenetic inhibition."*. In other words, optogenetic activation might inform what these populations *can* do, as opposed to what their role *is* under a normal, physiologically relevant, regime.

These observations, together with the fact that: 1) MSN exhibit low resting state firing rate, and 2) optogenetic activation of these cells has been shown to lead to lateral inhibition of other putative MSN units (Figure 2.16), suggest that inhibition experiments are a strong candidate for more readily interpretable results, regarding the functional role of direct and indirect pathways, going forwards.

3.3.2 On the role of indirect pathway in action suppression

When considering the mechanisms that give rise to adaptive action selection, considerably more attention has been paid to the positive aspect of this process, often neglecting the pivotal role action suppression plays. Several reasons could justify such a bias. For instance, while the "action" is often *observable*,

the corresponding suppressive process, if successful, will often remain latent and hidden from the experimenter's sight. Moreover, many common operant behaviors tested in a laboratory setting do not, *a priori*, require an explicit suppressive mechanism, and certainly not one under tight experimental control. As a result, in order to study the general process of action suppression, one ought to design paradigms that exaggerate the need for such a specific process and, finally, one should additionally be able to observe hallmarks of the reliance on such process (*e.g.,* in our task the fact that subjects often fail to remain immobile at the center port and report their choice in a way that is consistent with the dynamic reward contingencies of the task).

It is thus perhaps not surprising that, despite having long been implicated in action suppression (Mink & Thach, 1991; DeLong, 1990), very few studies to date have explicitly focused on the specific role indirect pathway plays during such a process. Instead, the vast majority of the literature focus on its role for the general process of action execution (Tai et al., 2012; Freeze et al., 2013; Tecuapetla et al., 2014, 2016). Nevertheless, previous key studies have reported findings consistent with the role of the indirect pathway supporting such a mechanism of action suppression that deserve mention.

In a series of experiments, Hikosaka and colleagues (Amita & Hikosaka, 2019; Hikosaka et al., 2019; H. F. Kim et al., 2017) trained monkeys in a task where subjects are sequentially presented with visual stimuli associated with either a large ("good target") or small/no reward ("bad target"). In order to maximize the total amount of collected reward, monkeys should refrain from saccading to "bad targets" and instead solely accept the offer associated with "good targets". After many sessions, subjects' behavior shows sensitivity to such contingencies. Consistent with the idea that the BG indirect pathway is involved in action suppression, in this case preventing animals from saccading to "bad" targets, the average activity of GPe neurons - that receive inhibitory input from iMSNs in the striatum - is significantly increased to "good" versus "bad targets". However, in some trials animals fail to reject the bad option. Strikingly, in this last fraction of trials, activity in the same cells was shown to be slightly higher than in correctly rejected trials, presumably due to lower activity levels in upstream striatal iMSNs.

The authors then proceed to show that, similarly to our task, such a pattern of activity is not only correlated with the failure to suppress a response but is instead necessary for it. Specifically, blocking GABA transmission in GPe - thus decoupling the effect iMSN activity has on GPe - renders animals unable to reject the "bad" option. These results are largely consistent with our physiology and manipulation data.

In Amita & Hikosaka (2019), authors focused on the overall change in firing rate associated with the ability of animals to successfully suppress a less than ideal action. Other studies, albeit leveraging different behavior paradigms, report that across individual cells, in different indirect pathway nuclei, response profiles are heterogeneous. Specifically, in both GPe (Yoshida & Tanaka, 2016; Mallet et al., 2016; Gu et al., 2020) and STN (Schmidt et al., 2013; Mosher et al., 2021), the authors found neurons that increased or decreased, respectively, their response during both movement execution and periods that require action suppression (be it proactive or reactive). How would one interpret these findings in light of the herein proposed role of the indirect pathway function?

One way to reconcile these two types of neuronal response profiles is to consider that the space of unsuitable, potentially competing, actions during a period of active immobility is likely different to the one during movement. In our task, for instance, a strong point can be made that the latent motor plan in need of suppressing is *"going short"* and *"going long"*, before and after the decision boundary, respectively. However, once such action is deployed, in order to generate the respective choice movement, the space of unsuitable actions might instead be enriched in motor programs that incompatible with the current movement or even perhaps other also considered, yet not taken, actions. This intuition can easily be thought of as a special case of the center-surround model presented in Chapter 1. During action suppression, the center of both the promoting and suppressing filter would perhaps be centered around the same action, whereas during movement, one would perhaps expect to find significantly more off-centered excitation of BG output in order to suppress other actions and allow the selected movement to be expressed.

Consistent with this general idea, we have also found some degree of heterogeneity in the responses of photo-identified indirect pathway medium spiny neurons. Briefly, by comparing the activity during *broken fixations* and time-matched control trials (Figure 2.14) we find that while some units did indeed decreased their firing rate a few hundred milliseconds prior to *breaking fixation* - consistent with the findings from Amita & Hikosaka (2019) in GPe - a second population of neurons showed the opposite dynamic and instead ramped up their activity immediately prior to movement onset. Critically, the degree to which iMSNs were negatively modulated during *broken fixations* was systematically related to how engaged they were during the period of action suppression. In other words, neurons that are engaged during the period of active immobility, shown to be causally involved in the ability of animals to suppress action (Figure 2.18), were the same neurons that show decreases in activity prior to *broken fixations*. Similarly, iMSNs that were suppressed during the delay period were those that showed a transient increase in activity aligned to the same event (Figure 2.15). Experiments combining simultaneous single-cell resolution calcium imaging and targeted optogenetic manipulation (Emiliani et al., 2015) could, in principle, dissect the role of these two populations in proactively supporting a tempting action and general movement, respectively.

Seemingly at odds with the interpretation that this latter population of iMSNs supported inhibition of other actions during movement, iMSNs optogenetic silencing did not impact the trajectory (Figure 2.3) nor the time mice took to report their choice at one of the side ports (Figure 2.19). From a technical point of view, it could be the case that our optogenetic inhibition protocol missed the period during which activity of this population was critical for the performance of the movement. Alternatively, it could also suggest a certain degree of *primacy* of the indirect pathway to suppress a prepotent action that must be eventually released, and less so to inhibit other actions during the execution of movement. Considering how small of an action-space animals are presented with in our task (e.g., "go left" or "go right"), it might result in overt actions that carry a large salience (*i.e., ballistic*). Interestingly enough, the effect of optogenetically silencing DLS iMSNs was the greatest during periods where the action was more

certain (*i.e., easy intervals*, Figure 2.20), suggesting that action suppression was indeed more important during periods where the drive to act was the greatest, which, intuitively, supports the idea that, in our task, a suppressive mechanism to prevent an action from being prematurely deployed takes precedence over a mechanism that simply increases the signal-to-noise when the same action is now being executed. If the hypothetical action promoting signal is inherently strong, disinhibition of other actions might turn out to be inconsequential. Different paradigms that systematically vary this *salience* parameter could present an opportunity to explicitly test this hypothesis.

The suppressive role of basal ganglia indirect pathway circuitry has been largely neglected especially when considering computational models of the BG. A meta-analysis of several such models (Helie et al., 2013) reported that while all considered instances included a direct pathway responsible for allowing actions to be expressed, only 7 out of 19 model instantiations included an indirect pathway explicitly. Given the obvious deficit in executive inhibitory control, be it motor or otherwise cognitive, that is often associated with BG lesions, the study of such a core hallmark of BG function must be prioritized. Unfortunately, the study of such a mechanism is often not trivial from a behavioral standpoint and carefully designed paradigms must be employed (Gomez-Marin et al., 2014; Krakauer et al., 2017; Niv, 2021) to isolate this process. Moving forward, if one of the main goals of systems and computational neuroscience is to understand cognition, I would argue that action suppression presents an obvious access point into what I would consider a cognitive process, since this inhibitory form of control must very often be covertly and internally generated, instead of being readily available from the immediate sensory vicinity of the animal.

3.3.3 Learning with two pathways

Adaptive behavior control must not only rely on hardwired and inflexible stimuli-response associations but instead must be able to dynamically learn what is the best action to perform in a particular context: a *policy*. Hence, any theory that then proposes BG as a potential neuronal substrate for action selection must also be able to layout how can BG learn said policies. In light of such a need, and

honoring the physiological knowledge of BG circuits, how might the suppressive dynamics observed in DLS during our task arise?

It has long been appreciated that dopamine can opposingly modulate the activity and plasticity of the two largest classes of medium spiny neurons. Specifically, increases and decreases in dopamine levels excite and induce long-term potentiation in dMSNs and iMSNs, respectively. Such physiological findings led to the idea dMSNs learn from events, and actions, that led to a better than expected outcome (positive reward prediction error, $\delta > 0$), while iMSNs appear to learn from worse-than-expected outcomes (negative reward prediction error, $\delta < 0$). This dichotomy can explain how the dynamics observed in our data arise and, more generally, how the indirect pathway circuitry might support adaptive action suppression (Collins & Frank, 2014; K. N. Gurney et al., 2015).

In our task, *breaking fixation* inevitably leads to a worse-than-expected outcome since animals will not be able to collect reward in that particular trial. Therefore, a premature response should always result in a negative RPE and a corresponding dip in dopamine levels. Such a dip might strengthen the synaptic weight between incoming contextual input (*e.g.,* time since first tone and/or action performed) and the corresponding indirect pathway medium spiny neurons. What would this mean for a behavior control policy? Taking into consideration the overall inhibitory influence the indirect pathway has on the output of BG, such a synaptic strengthening might result in a lower probability of performing the same action in the future and thus reducing the probability of *breaking fixation* in a subsequent trial. More generally, by combining the ability to both negatively modulate the probability of performing an action and learning from negative prediction errors, the indirect pathway is able to effectively guide behavior *away* from performing actions that resulted in poor outcomes in the past. Similarly, positing that the direct pathway learns from actions that led to a positive RPE in the past makes apparent how, during a period where moving never leads to reward, activity in the direct pathway is relatively lower.

3.3.4 Why two pathways?

Despite the elegance of such a conceptual framework, asymmetric learning in the two pathways has seen relatively few explicit implementations when modeling the dynamics of direct and indirect pathway medium spiny neuron responses, as well as behavior (but see Collins & Frank (2014); Delevich et al. (2020)). A potential reason for this seeming lack of interest is the apparent redundancy of having a rectified system that learns asymmetrically from positive and negative RPEs. Ultimately, the local policies derived from each of the two pathways must be combined - *e.g.,* summed - in order to generate a global behavior policy. Meanwhile, a simpler system endowed with a single pathway, able to bidirectionally regulate action probability and sensitive to the whole RPE range could, in principle, perform identically. The obvious question is then *why?*, why would the vertebrate brain converge to this two-pathway solution (Stephenson-Jones et al., 2011)?

Evolutionary constraints Let's begin with the usual culprit. As it was mentioned at the start of this chapter, circuits, and the computations that result from its organization, are not shaped by an intelligent designer, instead, evolution modifies, recycles and duplicates existing solutions that lead to a higher hill in some *fitness* landscape (Cisek, 2019). Therefore, while it is possible that an all-knowing engineer would come up with different, perhaps even superior, solutions for the problem of action selection in the vertebrate brain, splitting the control of BG output in two pathways was perhaps the readily found local maximum solution available throughout the evolutionary process. Why would this be a local optimum solution? One idea is that, similarly to one of the arguments given for the sparse coding found in the cortex (Simoncelli & Olshausen, 2001; Laughlin, 2001), having neurons that are generally silent, as is the case for striatal medium spiny neurons (Steiner & Tseng, 2016), might be energetically efficient. Conversely, if a single population was to control the output of BG, a higher baseline firing rate would probably be necessary in order to allow for bidirectional modulation. Moreover, early vertebrates had access to a very limited behavior repertoire, namely approaching and avoidance behaviors. It is possible that early instantiations of the direct and indirect pathway simply regulated the

interaction between these two behaviors (*e.g.,* promoting one and simultaneously inhibiting the other) and were later scaled to arbitrate between a greater number of actions.

Planning Another possible advantage introduced in Chapter 1 is the ability to *plan*. One of the most prevalent and useful skills animals have at their disposal is the ability to predict or anticipate future events based on previous experience (Gallistel, 1990). This ability affords animals time to prepare and pre-load a specific response to be expressed in a near future. A classic example of this kind of behavior is the systematic relationship between reaction times and *hazard* rate, or the probability of an event occurring given that it has not occurred yet, wherein animals display increasingly shorter reaction times for high likelihood events (Niemi & Näätänen, 1981; Mauk & Buonomano, 2004; Janssen & Shadlen, 2005), a phenomenon we were also able to observe in our task in both the hazard rate of *broken fixations* as a function of time (Figure 2.8) and in reaction times for completed trials as a function of interval presented (Figure 3.1). In other words, when provided with temporal information regarding the performance of an upcoming response, animals can, and will, leverage such information to prepare their movement. As it was stated before, this faculty allows animals to react much faster than they would otherwise and thus presents an obvious evolutionary advantage, especially in an ethologically relevant setting. Thus, having two pathways exerting distinct influence on the BG output might allow agents to simultaneously promote/select a future motor plan yet keep it "in check" until the time is ripe to deploy it. Moreover, if this suppressive signal shows some degree of action specificity, as we report in Chapter 2 when considering the unilateral inhibition and photometric recordings from the DLS indirect pathway, it could further be used to keep a specific action in check instead of simply increasing some sort of action threshold that might otherwise compromise other types of ongoing motor and cognitive function.

Modulate attitudes Encoding the value of *good* and *bad* options asymmetrically could potentially be leveraged to quickly modify a policy for behavior

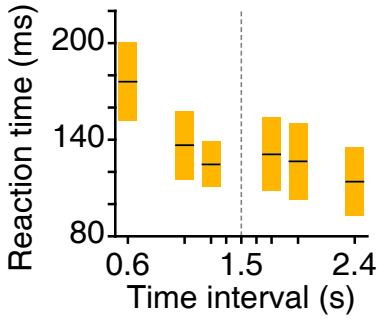


Figure 3.1. Mean of single animal median reaction times, as a function of presented interval, for a subset of animals used in the fiber photometry experiments. Bars represent mean and s.e.m.

control depending on, for example, the current motivational state of the animal. For instance, by increasing the relative *gain* of one pathway versus the other, one could adjust how much an agent should weigh *bad* versus *good* actions (Collins & Frank, 2014). Such a computation might be especially important in situations where an agent is not only paying attention to the long-term reward maximization but also to its safety (Garcia & Fernandez, 2015), or in situations where the agent must regulate exploring *versus* exploiting the current environment. Furthermore, since dopamine can not only support learning but also alter the relative excitability of direct and indirect pathways, one might conjecture that this neuromodulatory system could be able to implement such regulation of policy attitude. Indeed, while there is little doubt that dopamine encodes an RPE (Montague et al., 1996; Lau et al., 2017; H. R. Kim et al., 2020), the result of this computation could be leveraged to guide behavior in more ways than to just update the value of contexts and actions (Niv et al., 2007). It is thus not surprising that manipulation of the dopaminergic system, has been reported to alter risk attitudes (Tobler et al., 2007; Burke et al., 2018; Voon et al., 2011; Clark & Dagher, 2014; Beeler et al., 2010) and to regulate the explore-exploit decision-making process (Cinotti et al., 2019; Humphries et al., 2012; Chakroun et al., 2020; Beeler, 2012), even when accounting for potentially confounding learning effects(Beeler et al., 2010;

Tai et al., 2012). Unfortunately, the study of the impact dopamine has on risk attitudes might be further complicated by the observation that often, dopamine manipulations give rise to "U-shaped" effects (Beeler et al., 2010). Additional experiments will be needed to verify the exact nature of this interaction. Moreover, while dopamine presents itself as an obvious candidate to regulate the overall excitability of the two pathways, other neuromodulators, such as serotonin, could similarly regulate this process (Long et al., 2009; Rogers, 2011). Finally, it is unclear how this computation can be leveraged to further optimize adaptive behavior. Future efforts, especially theoretical, will likely provide critical insight into this question.

Mediation of parallel control modules Lastly, as it was brought up in the general introduction, the ability to promote and suppress might have arisen to meet the need to arbitrate among multiple parallel behavior controllers that seek partially incompatible access to actuators. Briefly, once parallel controllers are endowed with independent and unique representations of the world, local policies resulting from such unique views of the environment will likely be in conflict. Whether such an interaction is recurrent, hierarchical, or, most probably, something in between (*e.g.,* heterarchical), the ability to suppress, in addition to promote, seems critical in order to generate the best adaptive behavioral policy.

3.4 On the state and action spaces

Making choices in an adaptive way requires integrating multiple concurrent sources of information. These include limbic, cognitive and somatosensory signals that are combined in order to construct a representation of the *state* and *action* space of the world in which the biological agent inhabits. These are also the representations on top of which the relative advantage of behaviors can be learned, ultimately generating the behavior policy the biological agent follows. Ironically, one of the largest gaps in our understanding of BG function is the space in which context and actions are encoded.

Given the sparse input that arrives at the striatum from several different brain areas (Hunnicutt et al., 2016) it is not surprising that one can find MSN neuronal responses consistent with the representation of several different features of the environment and the animal. In CPu alone, a wealth of studies have observed neuronal correlates of low-level parameters of movement (*i.e.,* velocity, acceleration, joint angle, and kinematics (Barbera et al., 2016; Markowitz et al., 2018; Rueda-Orozco & Robbe, 2015; N. Kim et al., 2014; Jaeger et al., 1995)), sensory information (Reig & Silberberg, 2014; Sippy et al., 2015) , actions or sequences of actions (Jin & Costa, 2010; Sjöbom et al., 2020; Klaus et al., 2017; Markowitz et al., 2018), timing-related (Mello et al., 2015; Gouvêa et al., 2015) and spatial information (Hinman et al., 2019; van der Meer et al., 2010). Moreover, many of these correlates also multiplex information about the behavior context animals currently occupy. For example, neurons recorded from the caudate of primates during saccades show that, in addition to firing to movements, neuronal responses are also correlated to reward expectancy (Kawagoe et al., 2004; Lau & Glimcher, 2007), motivation/vigor (Fobbs et al., 2020) and, attention (Herman et al., 2020). In other words, consistent with the pivotal role BG plays in supporting learning, the striatal neurons exhibit responses that are consistent with the coding of the value of states and/or actions from which policies can be generated.

The abundance of distinct responses suggests that the striatum has access to a very rich and expressive set of basis functions that can be leveraged to represent the environment and actions. However, such variability also poses an enormous challenge when generating experimentally testable hypotheses regarding BG function. This problem justifiably warrants its own thesis but, for now, let us consider: how would one determine the *tuning* properties of a striatal neuron?

Paralleling the strategy used to map the receptive fields in early visual cortical areas (Hubel & Wiesel, 1962) one could, in theory, "sweep" across the state and action spaces while recording from MSNs. As the reader might imagine this is extremely difficult, if not entirely impossible. Not only is the dimensionality of this space vast, even before considering non-linear interactions between features, but, contrary to early visual areas, several of the dimensions to be probed are rarely under the explicit control of the experimenter (*e.g.,* actions). Two, almost

diametrically opposed, approaches have been adopted to tackle this problem. On one hand, by designing paradigms that isolate specific computations, researchers can parametrically probe the response of these neurons along some sub-dimension of the full feature space (*e.g.,* (Hikosaka et al., 2006, 2019; Lau & Glimcher, 2008; Gouvêa et al., 2015)). At the other extreme, one could leverage unconstrained behavior, followed by *post-hoc* detection of actions or syllables" (Klaus et al., 2017; Markowitz et al., 2018), and proceed to look for its putative neuronal correlates (Klaus et al., 2017). It is likely that this latter approach will be able to span a larger space than the former, however, due to the lack of experimental control, it becomes very difficult to probe dimensions not easily extracted from the low-level motor behavior features (*e.g.,* motivational state, value, context, etc), and it additionally rests on the assumption that any method to cluster behavior will be somewhat related to the encoding space of MSNs.

What do these challenges mean for the interpretation of "action specificity" we argue is present in our results? While it is true that we can't ascertain the granularity with which actions are encoded in medium spiny neurons in our task, it is nevertheless clear that we observe some degree of action specificity, namely across hemispheres, in the lateralized activity dynamics and the asymmetrical effects of unilateral optogenetics inhibition. Since DLS receives topographically organized input from the cortex, which in turn collects asymmetrical somatosensory information from the contralateral side of the body, it is probably fair to assume that MSNs from a specific hemisphere are involved in actions that recruit those same body parts and likely contain motor plans enriched in the contraversive direction. Moreover, decades of experiments have shown that both electrical and optogenetic activation of large striatal areas impacts contralateral actions the most (Nakamura & Hikosaka, 2006; Tai et al., 2012; Kravitz et al., 2010; Tecuapetla et al., 2014; J. Lee et al., 2020) and bulk activity in the striatum of a given hemisphere is systematically larger - for both direct and indirect pathway MSNs - during contraversive movements and stimuli (Cui et al., 2013; Tecuapetla et al., 2014).

It should be noted that while this line of argumentation is defensible for sensorimotor striatum, as one moves to areas enriched in more abstract state and

action spaces it is likely that such representations will become progressively less lateralized, or defined in a different set of coordinates altogether.

How can we increase the resolution with which we look at action? Future experiments that leverage high yield recording methods (single-cell calcium imaging or high-density electrophysiological recordings) and larger, relatively experimentally controlled, action spaces present a promising avenue. Meanwhile, AI in general, and deep neuronal networks in particular, might hold some of the answers. For instance, in Deep Q-Networks (DQN, (Mnih et al., 2015)), not unlike the mammalian cortex, initial layers are able to represent the state of the world through what is essentially a highly non-linear function approximator. Deeper layers use these representations to learn the value of different state-action pairs ($Q^{\pi}(s_t, a_t)$), reminiscent of what medium spiny neurons in the striatum might be learning. By training this network to perform tasks similar to those we employ in the lab one could, in theory, use them as an *in silico* model for the cortico-BG circuitry, for example, to find what kind of representations emerge in the network after training. Critically, as opposed to *in vivo* experiments, one could feasibly compute the full receptive field of any unit in the network by finding the inputs that maximize its activation (Yosinski et al., 2015). These methods could also be used to simulate "causal" experiments by externally modulating the activity of a specific sub-set of units and measuring the agent's behavior.

Whichever the path moving forward is, knowing the state and action space in which BG operates is, in my view, key to understand BG function.

3.5 On the parallel basal ganglia architecture and agent plurality

The basal ganglia are anatomically and functionally organized in largely independent parallel circuits. As cover in Chapter 1, a wealth of data suggests that each one of these loops honors the overall principles that govern BG anatomical and functional organization (DeLong, 1990) and might thus be implementing a similar computation. Despite potentially applying an identical function, the

inputs on which these BG circuits operate on are markedly distinct (Hunnicutt et al., 2014, 2016; Foster et al., 2021). These assumptions seed the core idea behind a multi-agent architecture of BG (Bornstein & Daw, 2011). From cortex alone, an hierarchy of representations is known to exist. In rodents, posterior, low-level, cortical areas carry somatosensory information about the immediate state of the world, while anterior areas are able to generate more abstract, multi-modal and temporal extended representations (Murray et al., 2014; Niv, 2019). As a result, striatal recipient areas of these inputs will be endowed with different representations on top of which learning can occur, and will possibly represent distinct behavior controllers with different "views" and "opinions" on the world (Bornstein & Daw, 2011; H. F. Kim & Hikosaka, 2015; Lau et al., 2017).

Several issues immediately arise when going from a single agent architecture (SAA) to a multi-agent one (MAA): What makes each agent unique? How are agents combined to generate a behavior policy? How do agents learn? Do agents share information among themselves? In this section I will give a brief account of my current view on these topics, how they relate to the model presented at the end of Chapter 2, and how future experimental and conceptual approaches might start to tackle some of these questions.

Apart from anatomical evidence, behavior and physiological data also seem to support the view that BG sub-circuits might work in parallel. For instance, in what is the classical textbook example, DLS has been shown to be necessary to acquire stimulus-response behavior contingencies (Yin et al., 2006, 2004), while DMS has been argued to support goal-directed learning (Shiflett et al., 2010; Balleine & O'Doherty, 2010; Yin et al., 2005). In early epochs of the learning process, DMS is preferentially recruited and later, as behavior progresses to become more habitual, DLS takes over (Goodman & Packard, 2016). Critically, experimental and theoretical evidence suggest that this recruitment is not sequential and, instead, both sub-circuits learn and compete for control from the get go (Bradfield & Balleine, 2013; Daw et al., 2005b). For instance, in a recent study, Bergstrom et al. (2018) show that, in a simple visual discrimination task, silencing rodent DLS early in training leads to faster learning when compared to control animals. If possessing multiple parallels controls leads, in some cases,

to suboptimal performance, why would evolution select such a BG feature? A common answer to this question involves considering that animals evolved to be able to learn to solve a vast variety of "tasks" (Daw et al., 2005b). The variability in the representations afforded to each sub-module will likely give rise to a continuous *adequateness* for each task, in which different agents will converge to different solutions and/or at different rates. Moreover, and closely related with the roles assigned to DLS and DMS during learning, efficient decision making will likely involve picking actions at the appropriate scale of abstraction. For example, early in learning, it might be advantageous to rely on previous models of the world to guide decision, hence engaging more associative BG loops. As learning progresses, a simpler association between stimulus and response might turn out to be more resource or time efficient, hence sensorimotor loops in the DLS might take over.

Similarly, at the end of Chapter 2, a MAA model is presented to explain the behavior in our task. The choice to model our results as a consequence of a MAA rested on two key observations. Firstly, as was mentioned before, from both physiological recordings and manipulation results, DLS seemed to be engaged to suppress and, not to promote action. Secondly, when trying to model our behavior with a SAA, endowed with a state-space representation containing information about elapsed time and second tone, such an instantiation would not produce *broken fixations*. Essentially, once the agent "knew" about the occurrence of the second tone, action values during the delay period would always be low and, thus, no action would be produced. Given such an observation, we added a second agent endowed with a slightly different state-space representation, one without access to second tone observations, able to generate the drive that eventually gave rise to *broken fixations*. Subsequent experiments revealed that, consistent with model predictions, DMS seems to be part of the circuit that implemented this second agent, responsible for carrying the drive to act throughout the delay period.

It is important to keep in mind that, while we were able to model the experimental results with the herein proposed state-space, other representations might be similarly successful. Namely, representations that give rise to a sort

of *predictive policy* might likewise end up generating premature behavior in our task. A possible general alternative going forward, is to model the non-DLS agent by integrating other known RL algorithms shown to generate a predictive-like state-space representation, such as *Successor Representation* (Dayan, 1993b; Stachenfeld et al., 2017). Meanwhile, confirming that a biological agent, and BG in particular, is indeed leveraging a specific state-space representation is a challenging endeavor. Nevertheless, a promising avenue would be to record the RPE, encoded in the dopaminergic input arriving at the striatum, and use it to constrain the space of possible, compatible, models (Motiwala et al., 2020).

Finally I would like to briefly touch upon the nature of the interaction between parallel agents. This is a broad and active topic of research (J. M. Smith, 1979; Prescott et al., 2006; Haruno & Kawato, 2006; Busoniu et al., 2008; Merel et al., 2019), and, while its full overview is far from the scope of this dissertation, some of its concepts are critical when trying to infer BG function in light of MAA.

In the herein proposed model, agents do not explicitly interact with one another. Instead, they become *aware* of each other's policy by interacting through behavior. For example, even though the non-DLS agent's policy pushes for an action during the delay period (*i.e.,* before the second tone), the DLS agent learns to suppress this policy by increasing the corresponding action values of the indirect pathway. Critically, the DLS agent is only informed of what action the general agent took (in this case, *breaking fixation* towards one of the sides) and the resulting, negative, RPE. Similarly to the state-space representation covered in the previous paragraph, it is likely that other forms of interaction between these two agents would yield similar behavior. However, with few exceptions, BG sub-modules are largely independent, hence a strong candidate substrate to support between agents communication seems lacking.

Nevertheless, other authors have pointed out how the few exceptions to the general anatomical segregation motif could actually allow for communication between sub-modules. One way this communication could take place is through cortical interactions that, given the reentrant nature of the circuit, might relay information across distinct BG loops (H. F. Kim & Hikosaka, 2015). Alterna-

tively, supported on evidence that RPEs made by limbic (VS) and associative (DMS) circuits are propagated to sensorimotor striatal circuits (DLS), some authors argue that some areas of the striatum can learn from errors provided from other loops (Haber et al., 2000; Haruno & Kawato, 2006). Finally, a recent study has also identified *open-loop* motifs in the BG circuitry. In Aoki et al. (2019), authors describe, in addition to the canonical sensorimotor loop, a second functional connection between ventral striatum and primary motor cortex. Future anatomical and physiological experiments will be critical to determine whether, and to what extent, communication between BG loops takes place and its functional relevance for adaptive behavior.

Stepping back, and abstracting from the precise details of the potential MAA implementation in BG, it might be interesting to consider in what ways such an architecture could be exploited to generate behavior[1]. Considering the aforementioned hierarchy of cortical representations, having different sub-modules might allow agents to learn independently and in a modular way. For instance, while the goal of a given sub-system might be to "make dinner", other lower-level systems might operate on simpler instructions such as "grab a knife" or even "activate muscle X". Accordingly, by using "local" errors instead of global ones, sub-agents can learn more efficiently. Meanwhile, since more abstract agents will likely have access to invariant context representations, their policies can be readily recycled across tasks. Additionally, such architectures are often easily scalable (and thus perhaps more likely to be selected by evolution) and also fault tolerant, in that removing one of the sub-modules will not lead to catastrophic failure (Kugler et al., 1990). Indeed, other RL algorithms, such as the *options framework* (Asada et al., 1996; Sutton et al., 1999), have sought to leverage this potential hierarchical nature of action planning.

For the sake of brevity much will be left to say about the exciting possibility that MAA represent yet another functional principle of BG organization. However, I'd like to finish this section by briefly considering how the field might proceed to study MAAs in the mammalian brain. For the first time since the dawn of Neuroscience, we are able to record from thousands of neurons simulta-

[1]Some of these ideas have been exquisitely well summarized in Merel et al. (2019)

neously (Steinmetz et al., 2019; Stringer et al., 2019) yet, when recording from all these neurons it is often the case that sensory, choice, and reward event responses are correlated across distant brain areas (Steinmetz et al., 2019). While it is certainly possible that all these variables are distributively coded and computed in the brain, it is also important to consider that the vast majority of paradigms we employ experimentally are relatively low dimensional when compared to the vast array of demands animals evolved to face in the wild. As a result, I would argue that a good place to start when studying MAA is *behavior*. In order to force decorrelations across agents, be it in state/action-space or in the policy, the behavior must be complex enough in order to allow for distinct agents to develop potentially different *views* and *opinions* of the world (Gomez-Marin et al., 2014; Krakauer et al., 2017; Lau et al., 2017). Walking down this path will probably be a hard and painstaking endeavor, and one that will likely involve not only relying on one paradigm but actually exposing animals to a multitude of different behavioral demands.

3.6 Outlook

Despite decades of accumulated knowledge on BG anatomy and physiology, a clear mechanistic understanding of its function is still lacking. One major point of debate, that we try to address in the work presented in this thesis, is whether or not the two major feedforward circuits of the BG, the direct and indirect pathways, functionally oppose each other during behavior. While we reproduced previous findings of co-activation during movement we also found that, consistent with classical models of BG function (Mink, 1996), during a period of action suppression one can observe signs of functional opponency between the two pathways in sensorimotor striatum. Moreover, we found that the indirect pathway is engaged during a period wherein animals must actively inhibit the urge to perform a tempting action. These findings add evidence towards a more complete understanding of BG function, and open exciting future research avenues.

Considering BG involvement in pathologies characterized by lack of inhibitory control (Penney & Young, 1983; Mink & Thach, 1993; Nigg, 2001; Qiu et al.,

2009), not only in the motor (*e.g.,* Huntington's disease) but also on more cognitive domains (*e.g.,* ADHD and OCD), it will be interesting to test whether the results we found in DLS translate to other striatal areas. For example, are persistent and intrusive thoughts, prevalent in many of these disorders, the outcome of a relatively lower indirect pathway activity in associative, or limbic, circuits? Such a research program is not only critical for the fundamental understanding of BG function, but also for developing new, and more effective, strategies for the treatment of patients ailed with these conditions.

One of the most exciting topics for future research, that is highlighted at the end of this thesis, is the multi-agent framework as a model for BG function. This framework is widespread in artificial intelligence (AI) and robotics where it found a large degree of success in problems requiring the control of autonomous agents Buşoniu et al. (2010); Knoblock (1990); Mussa-Ivaldi & Bizzi (2000); Todorov et al. (2005). However, it has received comparatively little attention in the neuroscience field, especially when modeling circuit function. I believe that much is to be gained from applying this framework to the brain. Firstly, both animals and artificial agents must solve similar problems in order to behave adaptively in their respective environments (Neftci & Averbeck, 2019). It should then stand to reason that solutions to AI problems might be useful to gain mechanistic insight and understanding of BG circuit function and vice-versa. Secondly, similarly to many MAAs found in artificial agents, brains are also hierarchically organized (Merel et al., 2019). While some circuits seem to specialize on *shallow*, sensory-motor processing common to many animal species, others produce abstractions that are increasingly *deeper*, complex, invariant and displaying a higher degree of temporal extendedness. Indeed, many of the abilities we hold to be specific to humans, might as well have arisen from progressively *deeper* and *deeper* layers of abstraction. Consistent with this idea, one of the main anatomical differences between primates (especially Humans) and other species is the large expansion of frontal cortical areas and, similarly, correspondingly added BG loops, where such *deeper* representations have been found before (Passingham, 1973; Deacon, 1998; Cisek, 2019). Lastly, if the past is any good predictor of the future, much

is to be gained from cross-pollination between AI and neuroscience in general, and BG research in particular (Montague et al., 1996).

Many questions will likely remain unanswered for decades to come regarding BG function. Given its critical involvement in normal motor behavior and cognition, its understanding will undoubtedly be crucial to infer overall vertebrate brain function. I, for one, cannot wait to see what the future holds.

References

Abdi, A., Mallet, N., Mohamed, F. Y., Sharott, A., Dodson, P. D., Nakamura, K. C., ... Magill, P. J. (2015, April). Prototypic and arkypallidal neurons in the dopamine-intact external globus pallidus. *J. Neurosci.*, *35*(17), 6667–6688.

Albin, R. L., Young, A. B., & Penney, J. B. (1989, October). The functional anatomy of basal ganglia disorders. *Trends Neurosci.*, *12*(10), 366–375.

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

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.*, *9*, 357–381.

Amita, H., & Hikosaka, O. (2019, August). Indirect pathway from caudate tail mediates rejection of bad objects in periphery. *Sci Adv*, *5*(8), eaaw9297.

Aoki, S., Smith, J. B., Li, H., Yan, X., Igarashi, M., Coulon, P., ... Jin, X. (2019, September). An open cortico-basal ganglia loop allows limbic control over motor output via the nigrothalamic pathway. *Elife*, *8*.

Aron, A. R. (2011, June). From reactive to proactive and selective control: Developing a richer model for stopping inappropriate responses. *Biol. Psychiatry*, *69*(12), e55–e68.

Aron, A. R., & Verbruggen, F. (2008, November). Stop the presses: Dissociating a selective from a global mechanism for stopping. *Psychol. Sci.*, *19*(11), 1146–1153.

Asada, M., Noda, S., Tawaratsumida, S., & Hosoda, K. (1996, May). Purposive behavior acquisition for a real robot by Vision-Based reinforcement learning. *Mach. Learn.*, *23*(2), 279–303.

Averbeck, B. B., Lehman, J., Jacobson, M., & Haber, S. N. (2014, July). Estimates of projection overlap and zones of convergence within frontal-striatal circuits. *J. Neurosci.*, *34*(29), 9497–9505.

Balleine, B. W., & O'Doherty, J. P. (2010, January). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, *35*(1), 48–69.

Barbeau, A. (1958, December). The understanding of involuntary movements: an historical approach. *J. Nerv. Ment. Dis.*, *127*(6), 469–489.

Barbera, G., Liang, B., Zhang, L., Gerfen, C. R., Culurciello, E., Chen, R., ... Lin, D.-T. (2016, October). Spatially compact neural clusters in the dorsal striatum encode locomotion relevant information. *Neuron*, *92*(1), 202–213.

Bar-Gad, I., Heimer, G., Ritov, Y., & Bergman, H. (2003, May). Functional correlations between neighboring neurons in the primate globus pallidus are weak or nonexistent. *J. Neurosci.*, *23*(10), 4012–4016.

Bar-Gad, I., Morris, G., & Bergman, H. (2003, December). Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Prog. Neurobiol.*, *71*(6), 439–473.

Barkley, R. A. (1997, January). Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychol. Bull.*, *121*(1), 65–94.

Baron, M. S., Wichmann, T., Ma, D., & DeLong, M. R. (2002, January). Effects of transient focal inactivation of the basal ganglia in parkinsonian primates. *J. Neurosci.*, *22*(2), 592–599.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear Mixed-Effects models using lme4. *Journal of Statistical Software, Articles*, *67*(1), 1–48.

Bechara, A., Noel, X., & Crone, E. A. (2006). Loss of willpower: Abnormal neural mechanisms of impulse control and decision making in addiction. *of implicit cognition and addiction.*

Beckstead, R. M., Domesick, V. B., & Nauta, W. J. (1979, October). Efferent connections of the substantia nigra and ventral tegmental area in the rat. *Brain Res.*, *175*(2), 191–217.

Beeler, J. A. (2012, August). Thorndike's law 2.0: Dopamine and the regulation of thrift. *Front. Neurosci.*, *6*, 116.

Beeler, J. A., Daw, N., Frazier, C. R. M., & Zhuang, X. (2010, November). Tonic dopamine modulates exploitation of reward learning. *Front. Behav. Neurosci.*, *4*, 170.

Benazzouz, A., Breit, S., Koudsie, A., Pollak, P., Krack, P., & Benabid, A.-L. (2002). Intraoperative microrecordings of the subthalamic nucleus in parkinson's disease. *Mov. Disord.*, *17 Suppl 3*, S145–9.

Benhamou, L., Kehat, O., & Cohen, D. (2014, February). Firing pattern characteristics of tonically active neurons in rat striatum: context dependent or species divergent? *J. Neurosci.*, *34*(6), 2299–2304.

Berendse, H. W., Galis-de Graaf, Y., & Groenewegen, H. J. (1992, February). Topographical organization and relationship with ventral striatal compartments of prefrontal corticostriatal projections in the rat. *J. Comp. Neurol.*, *316*(3), 314–347.

Bergman, T. J., & Beehner, J. C. (2021, December). Leveling with tinbergen: Four levels simplified to causes and consequences. *Evol. Anthropol.*.

Bergstrom, H. C., Lipkin, A. M., Lieberman, A. G., Pinard, C. R., Gunduz-Cinar, O., Brockway, E. T., ... Holmes, A. (2018, May). Dorsolateral striatum engagement interferes with early discrimination learning. *Cell Rep.*, *23*(8), 2264–2272.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004, March). Neural correlates of sensory and decision processes in auditory object identification. *Nat. Neurosci.*, *7*(3), 295–301.

Bolam, J. P., & Smith, Y. (1992, July). The striatum and the globus pallidus send convergent synaptic inputs onto single cells in the entopeduncular nucleus of the rat: a double anterograde labelling study combined with postembedding immunocytochemistry for GABA. *J. Comp. Neurol.*, *321*(3), 456–476.

Bolam, J. P., Smith, Y., Ingham, C. A., von Krosigk, M., & Smith, A. D. (1993). Convergence of synaptic terminals from the striatum and the globus pallidus onto single neurones in the substantia nigra and the entopeduncular nucleus. *Prog. Brain Res.*, *99*, 73–88.

Bolkan, S. S., Stone, I. R., Pinto, L., Ashwood, Z. C., Garcia, J. M. I., & others. (2021). Strong and opponent contributions of dorsomedial striatal pathways to behavior depends on cognitive demands and task strategy. *bioRxiv*.

Bornstein, A. M., & Daw, N. D. (2011, June). Multiplicity of control in the basal ganglia: computational roles of striatal subregions. *Curr. Opin. Neurobiol.*, *21*(3), 374–380.

Botvinick, M. M. (2007, December). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cogn. Affect. Behav. Neurosci.*, *7*(4), 356–366.

Box, G. E. P. (1976, December). Science and statistics. *J. Am. Stat. Assoc.*, *71*(356), 791–799.

Bradfield, L. A., & Balleine, B. W. (2013, January). Hierarchical and binary associations compete for behavioral control during instrumental biconditional discrimination. *J. Exp. Psychol. Anim. Behav. Process.*, *39*(1), 2–13.

Braitenberg, V. (1986). *Vehicles: Experiments in synthetic psychology*. MIT Press.

Braver, T. S. (2012, February). The variable nature of cognitive control: a dual mechanisms framework. *Trends Cogn. Sci.*, *16*(2), 106–113.

Breland, K., & Breland, M. (1961). The misbehavior of organisms. *Am. Psychol.*, *16*(11), 681–684.

Brimblecombe, K. R., & Cragg, S. J. (2017, February). The striosome and matrix compartments of the striatum: A path through the labyrinth from neurochemistry toward function. *ACS Chem. Neurosci.*, *8*(2), 235–242.

Broadbent, D. E. (1965, October). Information processing in the nervous system. *Science*, *150*(3695), 457–462.

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010, December). Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron*, *68*(5), 815–834.

Brooks, R. A. (1991). Intelligence without reason.

Burke, C. J., Soutschek, A., Weber, S., Raja Beharelle, A., Fehr, E., Haker, H., & Tobler, P. N. (2018, May). Dopamine Receptor-Specific contributions to the computation of value. *Neuropsychopharmacology*, *43*(6), 1415–1424.

Busemeyer, J. R., & Townsend, J. T. (1993, July). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychol. Rev.*, *100*(3), 432–459.

Busoniu, L., Babuska, R., & De Schutter, B. (2008). Multi-Agent reinforcement learning: An overview. *no.*, *2*, 156–172.

Buşoniu, L., Babuška, R., & De Schutter, B. (2010). *Multi-agent reinforcement learning: An overview*.

Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002, May). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.*, *26*(3), 321–352.

Carland, M. A., Thura, D., & Cisek, P. (2019, October). The urge to decide and act: Implications for brain function and dysfunction. *Neuroscientist*, *25*(5), 491–511.

Carlsson, A., Lindqvist, M., & Magnusson, T. (1957, November). 3,4-dihydroxyphenylalanine and 5-hydroxytryptophan as reserpine antagonists. *Nature*, *180*(4596), 1200.

Cartoni, E., Balleine, B., & Baldassarre, G. (2016, December). Appetitive pavlovian-instrumental transfer: A review. *Neurosci. Biobehav. Rev.*, *71*, 829–848.

Chakraborty, M., & Jarvis, E. D. (2015, December). Brain evolution by brain pathway duplication. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *370*(1684).

Chakroun, K., Mathar, D., Wiehler, A., Ganzer, F., & Peters, J. (2020, June). Dopaminergic modulation of the exploration/exploitation trade-off in human decision-making. *Elife*, *9*.

Chambers, C. D., Garavan, H., & Bellgrove, M. A. (2009, May). Insights into the neural basis of response inhibition from cognitive and clinical neuroscience. *Neurosci. Biobehav. Rev.*, *33*(5), 631–646.

Charnov, E. L. (1976, April). Optimal foraging, the marginal value theorem. *Theor. Popul. Biol.*, *9*(2), 129–136.

Chen, R., & Goldberg, J. H. (2020, December). Actor-critic reinforcement learning in the songbird. *Curr. Opin. Neurobiol.*, *65*, 1–9.

Chen, T.-W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., ... Kim, D. S. (2013, July). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, *499*(7458), 295–300.

Chen, W., de Hemptinne, C., Miller, A. M., Leibbrand, M., Little, S. J., Lim, D. A., ... Starr, P. A. (2020, May). Prefrontal-Subthalamic hyperdirect pathway modulates movement inhibition in humans. *Neuron*, *106*(4), 579–588.e3.

Chuong, A. S., Miri, M. L., Busskamp, V., Matthews, G. A. C., Acker, L. C., Sørensen, A. T., ... Boyden, E. S. (2014, August). Noninvasive optical inhibition with a red-shifted microbial rhodopsin. *Nat. Neurosci.*, *17*(8), 1123–1129.

Cinotti, F., Fresno, V., Aklil, N., Coutureau, E., Girard, B., Marchand, A. R., & Khamassi, M. (2019, May). Dopamine blockade impairs the exploration-exploitation trade-off in rats. *Sci. Rep.*, *9*(1), 6770.

Cisek, P. (2007, September). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *362*(1485), 1585–1599.

Cisek, P. (2019, October). Resynthesizing behavior through phylogenetic refinement. *Atten. Percept. Psychophys.*, *81*(7), 2265–2287.

Cisek, P., Puskas, G. A., & El-Murr, S. (2009, September). Decisions in changing conditions: the urgency-gating model. *J. Neurosci.*, *29*(37), 11560–11571.

Clark, C. A., & Dagher, A. (2014, May). The role of dopamine in risk taking: a specific look at parkinson's disease and gambling. *Front. Behav. Neurosci.*, *8*, 196.

Coddington, L. T., & Dudman, J. T. (2018, November). The timing of action determines reward prediction signals in identified midbrain dopamine neurons. *Nat. Neurosci.*, *21*(11), 1563–1573.

Coe, B. C., & Munoz, D. P. (2017, April). Mechanisms of saccade suppression revealed in the anti-saccade task. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *372*(1718).

Collins, A. G. E., & Frank, M. J. (2014, July). Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychol. Rev.*, *121*(3), 337–366.

Cox, J., & Witten, I. B. (2019, August). Striatal circuits for reward learning and decision-making. *Nat. Rev. Neurosci.*, *20*(8), 482–494.

Coxon, J. P., Stinear, C. M., & Byblow, W. D. (2007, March). Selective inhibition of movement. *J. Neurophysiol.*, *97*(3), 2480–2489.

Cruz, B. F., Soares, S., & Paton, J. J. (2020, July). *Dorsolateral striatal circuits support broadly opponent aspects of action suppression and production.*

Cubo, E., Shannon, K. M., Penn, R. D., & Kroin, J. S. (2000, November). Internal globus pallidotomy in dystonia secondary to huntington's disease. *Mov. Disord.*, *15*(6), 1248–1251.

Cui, G., Jun, S. B., Jin, X., Pham, M. D., Vogel, S. S., Lovinger, D. M., & Costa, R. M. (2013, February). Concurrent activation of striatal direct and indirect pathways during action initiation. *Nature*, *494*(7436), 238–242.

Dabney, W., Kurth-Nelson, Z., Uchida, N., Starkweather, C. K., Hassabis, D., Munos, R., & Botvinick, M. (2020, January). A distributional code for value in dopamine-based reinforcement learning. *Nature*, *577*(7792), 671–675.

Danjo, T., Yoshimi, K., Funabiki, K., Yawata, S., & Nakanishi, S. (2014, April). Aversive behavior induced by optogenetic inactivation of ventral tegmental area dopamine neurons is mediated by dopamine D2 receptors in the nucleus accumbens. *Proc. Natl. Acad. Sci. U. S. A.*, *111*(17), 6455–6460.

da Silva, J. A., Tecuapetla, F., Paixão, V., & Costa, R. M. (2018, February). Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature*, *554*(7691), 244–248.

Davies, N. B., Krebs, J. R., & West, S. A. (2012). *An introduction to behavioural ecology*. John Wiley & Sons.

Daw, N. D., Niv, Y., & Dayan, P. (2005a, December). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.*, *8*(12), 1704–1711.

Daw, N. D., Niv, Y., & Dayan, P. (2005b, December). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.*, *8*(12), 1704–1711.

Day, J. J., Roitman, M. F., Wightman, R. M., & Carelli, R. M. (2007, August). Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.*, *10*(8), 1020–1028.

Dayan, P. (1993a, July). Improving generalization for temporal difference learning: The successor representation. *Neural Comput.*, *5*(4), 613–624.

Dayan, P. (1993b, July). Improving generalization for temporal difference learning: The successor representation. *Neural Comput.*, *5*(4), 613–624.

Dayan, P., & Daw, N. D. (2008, December). Decision theory, reinforcement learning, and the brain. *Cogn. Affect. Behav. Neurosci.*, *8*(4), 429–453.

Dayan, P., Niv, Y., Seymour, B., & Daw, N. D. (2006, October). The misbehavior of value and the discipline of the will. *Neural Netw.*, *19*(8), 1153–1160.

Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain*. WW Norton & Company.

Delevich, K., Hoshal, B., Collins, A. G. E., & Wilbrecht, L. (2020). Choice suppression is achieved through opponent but not independent function of the striatal indirect pathway in mice.

DeLong, M. R. (1990, July). Primate models of movement disorders of basal ganglia origin. *Trends Neurosci.*, *13*(7), 281–285.

Deng, Y. P., Albin, R. L., Penney, J. B., Young, A. B., Anderson, K. D., & Reiner, A. (2004, June). Differential loss of striatal projection systems in huntington's disease: a quantitative immunohistochemical study. *J. Chem. Neuroanat.*, *27*(3), 143–164.

Deniau, J. M., & Chevalier, G. (1985, May). Disinhibition as a basic process in the expression of striatal functions. II. the striato-nigral influence on thalamocortical cells of the ventromedial thalamic nucleus. *Brain Res.*, *334*(2), 227–233.

Deniau, J. M., Menetrey, A., & Charpier, S. (1996, August). The lamellar organization of the rat substantia nigra pars reticulata: segregated patterns of striatal afferents and relationship to the topography of corticostriatal projections. *Neuroscience*, *73*(3), 761–781.

Denny-Brown, D., & Yanagisawa, N. (1976). The role of the basal ganglia in the initiation of movement. *Res. Publ. Assoc. Res. Nerv. Ment. Dis.*, *55*, 115–149.

Desmurget, M., & Turner, R. S. (2008, March). Testing basal ganglia motor functions through reversible inactivations in the posterior internal globus pallidus. *J. Neurophysiol.*, *99*(3), 1057–1076.

Dillon, D. G., & Pizzagalli, D. A. (2007, December). Inhibition of action, thought, and emotion: A selective neurobiological review. *Appl. Prev. Psychol.*, *12*(3), 99–114.

Dodson, P. D., Larvin, J. T., Duffell, J. M., Garas, F. N., Doig, N. M., Kessaris, N., ... Magill, P. J. (2015, April). Distinct developmental origins manifest in the specialized encoding of movement by adult neurons of the external globus pallidus. *Neuron*, *86*(2), 501–513.

Doll, B. B., Simon, D. A., & Daw, N. D. (2012, December). The ubiquity of model-based reinforcement learning. *Curr. Opin. Neurobiol.*, *22*(6), 1075–1081.

Dorfman, H. M., & Gershman, S. J. (2019, December). Controllability governs the balance between pavlovian and instrumental action selection. *Nat. Commun.*, *10*(1), 5826.

Doya, K. (1999). What are the computations of the cerebellum , the basal ganglia and the cerebral cortex ? , *12*, 961–974.

Dudman, J. T., & Krakauer, J. W. (2016, April). The basal ganglia: from motor commands to the control of vigor. *Curr. Opin. Neurobiol.*, *37*, 158–166.

Dunovan, K., Lynch, B., Molesworth, T., & Verstynen, T. (2015, September). Competing basal ganglia pathways determine the difference between stopping and deciding not to go. *Elife*, *4*, e08723.

Emiliani, V., Cohen, A. E., Deisseroth, K., & Häusser, M. (2015, October). All-Optical interrogation of neural circuits. *J. Neurosci.*, *35*(41), 13917–13926.

Engelhard, B., Finkelstein, J., Cox, J., Fleming, W., Jang, H. J., Ornelas, S., ... Witten, I. B. (2019, June). Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature*, *570*(7762), 509–513.

Eusebio, A., Pogosyan, A., Wang, S., Averbeck, B., Gaynor, L. D., Cantiniaux, S., ... Brown, P. (2009, August). Resonance in subthalamo-cortical circuits in parkinson's disease. *Brain*, *132*(Pt 8), 2139–2150.

Fink-Jensen, A., & Mikkelsen, J. D. (1991, February). A direct neuronal projection from the entopeduncular nucleus to the globus pallidus. a PHA-L anterograde tracing study in the rat. *Brain Res.*, *542*(1), 175–179.

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003, March). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, *299*(5614), 1898–1902.

Flaherty, A. W., & Graybiel, A. M. (1993, August). Output architecture of the primate putamen. *J. Neurosci.*, *13*(8), 3222–3237.

Fobbs, W. C., Bariselli, S., Licholai, J. A., Miyazaki, N. L., Matikainen-Ankney, B. A., Creed, M. C., & Kravitz, A. V. (2020, February). Continuous representations of speed by striatal medium spiny neurons. *J. Neurosci.*, *40*(8), 1679–1688.

Ford, K. A., & Everling, S. (2009, October). Neural activity in primate caudate nucleus associated with pro- and antisaccades. *J. Neurophysiol.*, *102*(4), 2334–2341.

Foster, N. N., Barry, J., Korobkova, L., Garcia, L., Gao, L., Becerra, M., ... Dong, H.-W. (2021, October). The mouse cortico-basal ganglia-thalamic network. *Nature*, *598*(7879), 188–194.

Frank, M. J., Seeberger, L. C., & O'reilly, R. C. (2004, December). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, *306*(5703), 1940–1943.

Franklin, K. B. J., & Paxinos, G. (2008). *The mouse brain in stereotaxic coordinates 3rd edn.* Academic press.

Freeze, B. S., Kravitz, A. V., Hammack, N., Berke, J. D., & Kreitzer, A. C. (2013, November). Control of basal ganglia output by direct and indirect pathway projection neurons. *J. Neurosci.*, *33*(47), 18531–18539.

Fujiyama, F., Sohn, J., Nakano, T., Furuta, T., Nakamura, K. C., Matsuda, W., & Kaneko, T. (2011, February). Exclusive and common targets of neostriatofugal projections of rat striosome neurons: a single neuron-tracing study using a viral vector. *Eur. J. Neurosci.*, *33*(4), 668–677.

Gadagkar, V., Puzerey, P. A., Chen, R., Baird-Daniel, E., Farhang, A. R., & Goldberg, J. H. (2016, December). Dopamine neurons encode performance error in singing birds. *Science*, *354*(6317), 1278–1282.

Gallistel, C. R. (1990). The organization of learning. *Learning, development, and conceptual change.*, *648*.

Gan, J. O., Walton, M. E., & Phillips, P. E. M. (2010, January). Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. *Nat. Neurosci.*, *13*(1), 25–27.

Garcia, J., & Fernandez, F. (2015). *A comprehensive survey on safe reinforcement learning.* `https://www.jmlr.org/papers/volume16/garcia15a/garcia15a.pdf`. (Accessed: 2021-12-13)

Geddes, C. E., Li, H., & Jin, X. (2018, June). Optogenetic editing reveals the hierarchical organization of learned action sequences. *Cell*, *174*(1), 32–43.e15.

Gerfen, C. R. (1984). The neostriatal mosaic: compartmentalization of corticostriatal input and striatonigral output systems. *Nature*, *311*(5985), 461–464.

Gerfen, C. R. (1992, April). The neostriatal mosaic: multiple levels of compartmental organization. *Trends Neurosci.*, *15*(4), 133–139.

Gerfen, C. R., Paletzki, R., & Heintz, N. (2013, December). GENSAT BAC cre-recombinase driver lines to study the functional organization of cerebral cortical and basal ganglia circuits. *Neuron*, *80*(6), 1368–1383.

Gerfen, C. R., & Surmeier, D. J. (2011). Modulation of striatal projection systems by dopamine. *Annu. Rev. Neurosci.*, *34*, 441–466.

Gibbon, J. (1977, May). Scalar expectancy theory and weber's law in animal timing. *Psychol. Rev.*, *84*(3), 279–325.

Gomez-Marin, A., Paton, J. J., Kampff, A. R., Costa, R. M., & Mainen, Z. F. (2014, November). Big behavioral data: psychology, ethology and the foundations of neuroscience. *Nat. Neurosci.*, *17*(11), 1455–1462.

Goodman, J., & Packard, M. G. (2016, January). Chapter 35 - memory systems of the basal ganglia. In H. Steiner & K. Y. Tseng (Eds.), *Handbook of behavioral neuroscience* (Vol. 24, pp. 725–740). Elsevier.

Gouvêa, T. S., Monteiro, T., Motiwala, A., Soares, S., Machens, C. K., & Paton, J. J. (2015, June). *Striatal dynamics explain duration judgments* (Tech. Rep.).

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.*, *31*, 359–387.

Graybiel, A. M., & Ragsdale, C. W., Jr. (1978, November). Histochemically distinct compartments in the striatum of human, monkeys, and cat demonstrated by acetylthiocholinesterase staining. *Proc. Natl. Acad. Sci. U. S. A.*, *75*(11), 5723–5726.

Gremel, C. M., & Costa, R. M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nat. Commun.*, *4*, 2264.

Grillner, S., Hellgren, J., Ménard, A., Saitoh, K., & Wikström, M. A. (2005, July). Mechanisms for selection of basic motor programs–roles for the striatum and pallidum. *Trends Neurosci.*, *28*(7), 364–370.

Grillner, S., & Robertson, B. (2016, October). The basal ganglia over 500 million years. *Curr. Biol.*, *26*(20), R1088–R1100.

Gross, C. G. (2007, July). The discovery of motor cortex and its background. *J. Hist. Neurosci.*, *16*(3), 320–331.

Gu, B.-M., Schmidt, R., & Berke, J. D. (2020, June). Globus pallidus dynamics reveal covert strategies for behavioral inhibition. *Elife*, *9*.

Guitart-Masip, M., Duzel, E., Dolan, R., & Dayan, P. (2014, April). Action versus valence in decision making. *Trends Cogn. Sci.*, *18*(4), 194–202.

Gulley, J. M., Kuwajima, M., Mayhill, E., & Rebec, G. V. (1999, October). Behavior-related changes in the activity of substantia nigra pars reticulata neurons in freely moving rats. *Brain Res.*, *845*(1), 68–76.

Gurney, K., Prescott, T. J., & Redgrave, P. (1998, October). The basal ganglia viewed as an action selection device. In *International conference on artificial neural networks* (pp. 1033–1038). unknown.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001, June). A computational model of action selection in the basal ganglia. II. analysis and simulation of behaviour. *Biol. Cybern.*, *84*(6), 411–423.

Gurney, K. N., Humphries, M. D., & Redgrave, P. (2015, January). A new framework for cortico-striatal plasticity: behavioural theory meets in vitro data at the reinforcement-action interface. *PLoS Biol.*, *13*(1), e1002034.

Haber, S. N. (2003, December). The primate basal ganglia: parallel and integrative networks. *J. Chem. Neuroanat.*, *26*(4), 317–330.

Haber, S. N. (2014, December). The place of dopamine in the cortico-basal ganglia circuit. *Neuroscience*, *282*, 248–257.

Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000, March). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J. Neurosci.*, *20*(6), 2369–2382.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Res.*, *18*(10), 1279–1296.

Hamid, A. A., Frank, M. J., & Moore, C. I. (2021, May). Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell*, *184*(10), 2733–2749.e16.

Han, X., Chow, B. Y., Zhou, H., Klapoetke, N. C., Chuong, A., Rajimehr, R., ... Boyden, E. S. (2011, April). A high-light sensitivity optical neural silencer: development and application to optogenetic control of non-human primate cortex. *Front. Syst. Neurosci.*, *5*, 18.

Haruno, M., & Kawato, M. (2006, October). Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. *Neural Netw.*, *19*(8), 1242–1254.

Hawley, J. S., & Weiner, W. J. (2012, February). Hemiballismus: current concepts and review. *Parkinsonism Relat. Disord.*, *18*(2), 125–129.

Hazrati, L. N., & Parent, A. (1992a, December). Differential patterns of arborization of striatal and subthalamic fibers in the two pallidal segments in primates. *Brain Res.*, *598*(1-2), 311–315.

Hazrati, L. N., & Parent, A. (1992b, October). The striatopallidal projection displays a high degree of anatomical specificity in the primate. *Brain Res.*, *592*(1-2), 213–227.

Hazrati, L. N., Parent, A., Mitchell, S., & Haber, S. N. (1990, November). Evidence for interconnections between the two segments of the globus pallidus in primates: a PHA-L anterograde tracing study. *Brain Res.*, *533*(1), 171–175.

Hegeman, D. J., Hong, E. S., Hernández, V. M., & Chan, C. S. (2016, May). The external globus pallidus: progress and perspectives. *Eur. J. Neurosci.*, *43*(10), 1239–1265.

Helie, S., Chakravarthy, S., & Moustafa, A. A. (2013, December). Exploring the cognitive and motor functions of the basal ganglia: an integrative review of computational cognitive neuroscience models. *Front. Comput. Neurosci.*, *7*, 174.

Herman, J. P., Arcizet, F., & Krauzlis, R. J. (2020, September). Attention-related modulation of caudate neurons depends on superior colliculus activity. *Elife*, *9*.

Hernández, V. M., Hegeman, D. J., Cui, Q., Kelver, D. A., Fiske, M. P., Glajch, K. E., ... Chan, C. S. (2015, August). Parvalbumin+ neurons and npas1+ neurons are distinct neuron classes in the mouse external globus pallidus. *J. Neurosci.*, *35*(34), 11830–11847.

Hernández-López, S., Bargas, J., Surmeier, D. J., Reyes, A., & Galarraga, E. (1997, May). D1 receptor activation enhances evoked discharge in neostriatal medium spiny neurons by modulating an l-type ca2+ conductance. *J. Neurosci.*, *17*(9), 3334–3342.

Hernandez-Lopez, S., Tkatch, T., Perez-Garci, E., Galarraga, E., Bargas, J., Hamm, H., & Surmeier, D. J. (2000, December). D2 dopamine receptors in striatal medium spiny neurons reduce l-type ca2+ currents and excitability via a novel PLC[beta]1-IP3-calcineurin-signaling cascade. *J. Neurosci.*, *20*(24), 8987–8995.

Herrnstein, R. J. (1961, July). Relative and absolute strength of response as a function of frequency of reinforcement. *J. Exp. Anal. Behav.*, *4*, 267–272.

Hikosaka, O., Kim, H. F., Amita, H., Yasuda, M., Isoda, M., Tachibana, Y., & Yoshida, A. (2019, March). Direct and indirect pathways for choosing objects and actions. *Eur. J. Neurosci.*, *49*(5), 637–645.

Hikosaka, O., Nakamura, K., & Nakahara, H. (2006, February). Basal ganglia orient eyes to reward. *J. Neurophysiol.*, *95*(2), 567–584.

Hikosaka, O., Takikawa, Y., & Kawagoe, R. (2000, July). Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiol. Rev.*, *80*(3), 953–978.

Hikosaka, O., & Wurtz, R. H. (1985, January). Modification of saccadic eye movements by GABA-related substances. II. effects of muscimol in monkey substantia nigra pars reticulata. *J. Neurophysiol.*, *53*(1), 292–308.

Hinman, J. R., Chapman, G. W., & Hasselmo, M. E. (2019, June). Neuronal representation of environmental boundaries in egocentric coordinates. *Nat. Commun.*, *10*(1), 2772.

Hintiryan, H., Foster, N. N., Bowman, I., Bay, M., Song, M. Y., Gou, L., ... Dong, H.-W. (2016, August). The mouse cortico-striatal projectome. *Nat. Neurosci.*, *19*(8), 1100–1114.

Houk, J. C., Adams, J. L., & Barto, A. G. (1994). A model of how the basal ganglia might generate and use neural signals that predict reinforcement. In *Workshop paper* (Vol. 13). unknown.

Howe, M. W., & Dombeck, D. A. (2016, July). Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature*, *535*(7613), 505–510.

Hubel, D. H., & Wiesel, T. N. (1962, January). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, *160*, 106–154.

Humphries, M. D., Khamassi, M., & Gurney, K. (2012, February). Dopaminergic control of the Exploration-Exploitation Trade-Off via the basal ganglia. *Front. Neurosci.*, *6*, 9.

Hunnicutt, B. J., Jongbloets, B. C., Birdsong, W. T., Gertz, K. J., Zhong, H., & Mao, T. (2016, November). A comprehensive excitatory input map of the striatum reveals novel functional organization. *Elife*, *5*(vember2016), 1–32.

Hunnicutt, B. J., Long, B. R., Kusefoglu, D., Gertz, K. J., Zhong, H., & Mao, T. (2014, September). A comprehensive thalamocortical projection map at the mesoscopic level. *Nat. Neurosci.*, *17*(9), 1276–1285.

Hutchison, W. D., Allan, R. J., Opitz, H., Levy, R., Dostrovsky, J. O., Lang, A. E., & Lozano, A. M. (1998, October). Neurophysiological identification of the subthalamic nucleus in surgery for parkinson's disease. *Ann. Neurol.*, *44*(4), 622–628.

Iino, Y., Sawada, T., Yamaguchi, K., Tajiri, M., Ishii, S., Kasai, H., & Yagishita, S. (2020, March). Dopamine D2 receptors in discrimination learning and spine enlargement. *Nature*, *579*(7800), 555–560.

J. J. A. van Iersel, & A. C. Angela Bol. (1958). Preening of two tern species. a study on displacement activities. *Behaviour*, *13*(1/2), 1–88.

Jaeger, D., Gilman, S., & Aldridge, J. W. (1995, October). Neuronal activity in the striatum and pallidum of primates related to the execution of externally cued reaching movements. *Brain Res.*, *694*(1-2), 111–127.

Jaeger, D., Kita, H., & Wilson, C. J. (1994, November). Surround inhibition among projection neurons is weak or nonexistent in the rat neostriatum. *J. Neurophysiol.*, *72*(5), 2555–2558.

Janssen, P., & Shadlen, M. N. (2005, February). A representation of the hazard rate of elapsed time in macaque area LIP. *Nat. Neurosci.*, *8*(2), 234–241.

Jin, X., & Costa, R. M. (2010, July). Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature*, *466*(7305), 457–462.

Joel, D., Niv, Y., & Ruppin, E. (2002, June). Actor–critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Netw.*, *15*(4), 535–547.

Joel, D., & Weiner, I. (1994, November). The organization of the basal ganglia-thalamocortical circuits: open interconnected rather than closed segregated. *Neuroscience*, *63*(2), 363–379.

Joel, D., & Weiner, I. (2000). The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum. *Neuroscience*, *96*(3), 451–474.

Kawagoe, R., Takikawa, Y., & Hikosaka, O. (2004, February). Reward-predicting activity of dopamine and caudate neurons–a possible mechanism of motivational control of saccadic eye movement. *J. Neurophysiol.*, *91*(2), 1013–1024.

Kemp, J. M., & Powell, T. P. (1970). The cortico-striate projection in the monkey. *Brain*, *93*(3), 525–546.

Kemp, J. M., & Powell, T. P. (1971, September). The connexions of the striatum and globus pallidus: synthesis and speculation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *262*(845), 441–457.

Kim, H. F., Amita, H., & Hikosaka, O. (2017, May). Indirect pathway of caudal basal ganglia for rejection of valueless visual objects. *Neuron*, *94*(4), 920–930.e3.

Kim, H. F., & Hikosaka, O. (2013, September). Distinct basal ganglia circuits controlling behaviors guided by flexible and stable values. *Neuron*, *79*(5), 1001–1010.

Kim, H. F., & Hikosaka, O. (2015, July). Parallel basal ganglia circuits for voluntary and automatic behaviour to reach rewards. *Brain*, *138*(Pt 7), 1776–1800.

Kim, H. R., Malik, A. N., Mikhael, J. G., Bech, P., Tsutsui-Kimura, I., Sun, F., ... Uchida, N. (2020, December). A unified framework for dopamine signals across timescales. *Cell*, *183*(6), 1600–1616.e25.

Kim, N., Barter, J. W., Sukharnikova, T., & Yin, H. H. (2014, November). Striatal firing rate reflects head movement velocity. *Eur. J. Neurosci.*, *40*(10), 3481–3490.

Kincaid, A. E., Zheng, T., & Wilson, C. J. (1998, June). Connectivity and convergence of single corticostriatal axons. *J. Neurosci.*, *18*(12), 4722–4731.

Kita, H., & Kita, T. (2011, July). Cortical stimulation evokes abnormal responses in the dopamine-depleted rat basal ganglia. *J. Neurosci.*, *31*(28), 10311–10322.

Kita, H., & Kitai, S. T. (1994, February). The morphology of globus pallidus projection neurons in the rat: an intracellular staining study. *Brain Res.*, *636*(2), 308–319.

Kita, H., Tokuno, H., & Nambu, A. (1999, May). Monkey globus pallidus external segment neurons projecting to the neostriatum. *Neuroreport*, *10*(7), 1467–1472.

Klaus, A., Martins, G. J., Paixao, V. B., Zhou, P., Paninski, L., & Costa, R. M. (2017, November). The spatiotemporal organization of the striatum encodes action space. *Neuron*, *96*(4), 1171–1180.e7.

Knoblock, C. A. (1990). Learning abstraction hierarchies for problem solving. *AAAI*.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996, September). A neostriatal habit learning system in humans. *Science*, *273*(5280), 1399–1402.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017, February). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron*, *93*(3), 480–490.

Kravitz, A. V., Freeze, B. S., Parker, P. R. L., Kay, K., Thwin, M. T., Deisseroth, K., & Kreitzer, A. C. (2010, July). Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature*, *466*(7306), 622–626.

Kravitz, A. V., Tye, L. D., & Kreitzer, A. C. (2012, June). Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nat. Neurosci.*, *15*(6), 816–818.

Kreitzer, A. C., & Malenka, R. C. (2008, November). Striatal plasticity and basal ganglia circuit function. *Neuron*, *60*(4), 543–554.

Kuffler, S. W. (1953, January). Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.*, *16*(1), 37–68.

Kugler, P. N., Shaw, R. E., Vincente, K. J., & Kinsella-Shaw, J. (1990). Inquiry into intentional systems i: Issues in ecological physics. *Psychol. Res.*, *52*(2-3), 98–121.

Kvitsiani, D., Ranade, S., Hangya, B., Taniguchi, H., Huang, J. Z., & Kepecs, A. (2013, June). Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature*, *498*(7454), 363–366.

Lahiri, A. K., & Bevan, M. D. (2020, April). Dopaminergic transmission rapidly and persistently enhances excitability of D1 Receptor-Expressing striatal projection neurons. *Neuron*, *106*(2), 277–290.e6.

Lanciego, J. L., Luquin, N., & Obeso, J. A. (2012, December). Functional neuroanatomy of the basal ganglia. *Cold Spring Harb. Perspect. Med.*, *2*(12), a009621.

Lanska, D. J. (2009, January). Chapter 33 the history of movement disorders. In M. J. Aminoff, F. Boller, & D. F. Swaab (Eds.), *Handbook of clinical neurology* (Vol. 95, pp. 501–546). Elsevier.

Lau, B., & Glimcher, P. W. (2007, December). Action and outcome encoding in the primate caudate nucleus. *J. Neurosci.*, *27*(52), 14502–14514.

Lau, B., & Glimcher, P. W. (2008, May). Value representations in the primate striatum during matching behavior. *Neuron*, *58*(3), 451–463.

Lau, B., Monteiro, T., & Paton, J. J. (2017, October). The many worlds hypothesis of dopamine prediction error: implications of a parallel circuit architecture in the basal ganglia. *Curr. Opin. Neurobiol.*, *46*, 241–247.

Laughlin, S. B. (2001, August). Energy as a constraint on the coding and processing of sensory information. *Curr. Opin. Neurobiol.*, *11*(4), 475–480.

Lee, J., Wang, W., & Sabatini, B. L. (2020, November). Anatomically segregated basal ganglia pathways allow parallel behavioral modulation. *Nat. Neurosci.*, *23*(11), 1388–1398.

Lee, S. J., Lodder, B., Chen, Y., Patriarchi, T., Tian, L., & Sabatini, B. L. (2021, February). Cell-type-specific asynchronous modulation of PKA by dopamine in learning. *Nature*, *590*(7846), 451–456.

Lenth, R. (2016). Least-Squares means: The R package lsmeans. *Journal of Statistical Software, Articles*, *69*(1), 1–33.

Levy, R., Dostrovsky, J. O., Lang, A. E., Sime, E., Hutchison, W. D., & Lozano, A. M. (2001, July). Effects of apomorphine on subthalamic nucleus and globus pallidus internus neurons in patients with parkinson's disease. *J. Neurophysiol.*, *86*(1), 249–260.

Levy, R., Lang, A. E., Dostrovsky, J. O., Pahapill, P., Romas, J., Saint-Cyr, J., ... Lozano, A. M. (2001, October). Lidocaine and muscimol microinjections in subthalamic nucleus reverse parkinsonian symptoms. *Brain*, *124*(Pt 10), 2105–2118.

Lima, S. Q., Hromádka, T., Znamenskiy, P., & Zador, A. M. (2009, July). PINP: a new method of tagging neuronal populations for identification during in vivo electrophysiological recording. *PLoS One*, *4*(7), e6099.

Litvak, V., Jha, A., Eusebio, A., Oostenveld, R., Foltynie, T., Limousin, P., ... Brown, P. (2011, February). Resting oscillatory cortico-subthalamic connectivity in patients with parkinson's disease. *Brain*, *134*(Pt 2), 359–374.

Loas, G., & Krystkowiak, P. (2015). Is state anhedonia characteristic of parkinson's disease? *Adv. Aging Res.*, *04*(06), 225–229.

Logan, G. D. (1981). Attention, automaticity, and the ability to stop a speeded choice response. *Attention and performance IX*, 205–222.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychol. Rev.*, *91*(3), 295–327.

Long, A. B., Kuhn, C. M., & Platt, M. L. (2009, December). Serotonin shapes risky decision making in monkeys. *Soc. Cogn. Affect. Neurosci.*, *4*(4), 346–356.

Lopes, G., Bonacchi, N., Frazão, J., Neto, J. P., Atallah, B. V., Soares, S., ... Kampff, A. R. (2015, April). Bonsai: an event-based framework for processing and controlling data streams. *Front. Neuroinform.*, *9*, 7.

Lozano, A. M., Lang, A. E., Levy, R., Hutchison, W., & Dostrovsky, J. (2000, April). Neuronal recordings in parkinson's disease patients with dyskinesias induced by apomorphine. *Ann. Neurol.*, *47*(4 Suppl 1), S141–6.

Lu, M. T., Preston, J. B., & Strick, P. L. (1994, March). Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. *J. Comp. Neurol.*, *341*(3), 375–392.

MacDonald, M. E., Ambrose, C. M., & Duyao MP, L. C. S. L. B. G. T. S. J. M. G. N. M. H., Myers RH. (1993, March). A novel gene containing a trinucleotide repeat that is expanded and unstable on huntington's disease chromosomes. the huntington's disease collaborative research group. *Cell*, *72*(6), 971–983.

Madisen, L., Mao, T., Koch, H., Zhuo, J.-M., Berenyi, A., Fujisawa, S., ... Zeng, H. (2012, March). A toolbox of cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nat. Neurosci.*, *15*(5), 793–802.

Maia, T. V. (2009, December). Reinforcement learning, conditioning, and the brain: Successes and challenges. *Cogn. Affect. Behav. Neurosci.*, *9*(4), 343–364.

Majid, D. S. A., Cai, W., Corey-Bloom, J., & Aron, A. R. (2013, August). Proactive selective response suppression is implemented via the basal ganglia. *J. Neurosci.*, *33*(33), 13259–13269.

Mallet, N., Ballion, B., Le Moine, C., & Gonon, F. (2006, April). Cortical inputs and GABA interneurons imbalance projection neurons in the striatum of parkinsonian rats. *J. Neurosci.*, *26*(14), 3875–3884.

Mallet, N., Leblois, A., Maurice, N., & Beurrier, C. (2019, December). Striatal cholinergic interneurons: How to elucidate their function in health and disease. *Front. Pharmacol.*, *10*, 1488.

Mallet, N., Micklem, B. R., Henny, P., Brown, M. T., Williams, C., Bolam, J. P., ... Magill, P. J. (2012, June). Dichotomous organization of the external globus pallidus. *Neuron*, *74*(6), 1075–1086.

Mallet, N., Schmidt, R., Leventhal, D., Chen, F., Amer, N., Boraud, T., & Berke, J. D. (2016, January). Arkypallidal cells send a stop signal to striatum. *Neuron*, *89*(2), 308–316.

Maltenfort, M. G., Heckman, C. J., & Rymer, W. Z. (1998, July). Decorrelating actions of renshaw interneurons on the firing of spinal motoneurons within a motor nucleus: a simulation study. *J. Neurophysiol.*, *80*(1), 309–323.

Markowitz, J. E., Gillis, W. F., Beron, C. C., Neufeld, S. Q., Robertson, K., Bhagat, N. D., ... Datta, S. R. (2018, June). The striatum organizes 3D behavior via Moment-to-Moment action selection. *Cell*, *174*(1), 44–58.e17.

Marsh, B. (2002). Do animals use heuristics? *Journal o fBioeconomics*, *4*, 49–56.

Martin, J. P., & Alcock, N. S. (1934, December). HEMICHOREA ASSOCIATED WITH a LESION OF THE CORPUS LUYSII. *Brain*, *57*(4), 504–516.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018, September). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.*, *21*(9), 1281–1289.

Matias, S., Lottem, E., Dugué, G. P., & Mainen, Z. F. (2017, March). Activity patterns of serotonin neurons underlying cognitive flexibility. *Elife*, *6*.

Mauk, M. D., & Buonomano, D. V. (2004). The neural basis of temporal processing. *Annu. Rev. Neurosci.*, *27*(1), 307–340.

McDowell, J. J. (2005, July). On the classic and modern theories of matching. *J. Exp. Anal. Behav.*, *84*(1), 111–127.

McFarland, D. (1989). *Problems of animal behaviour*. Longman Publishing Group.

McFarland, D. J. (1969, May). Mechanisms of behavioural disinhibition. *Anim. Behav.*, *17*, 238–242.

Mello, G. B. M., Soares, S., & Paton, J. J. (2015, May). A scalable population code for time in the striatum. *Curr. Biol.*, *25*(9), 1113–1122.

Menegas, W., Akiti, K., Amo, R., Uchida, N., & Watabe-Uchida, M. (2018, October). Dopamine neurons projecting to the posterior striatum reinforce avoidance of threatening stimuli. *Nat. Neurosci.*, *21*(10), 1421–1430.

Merel, J., Botvinick, M., & Wayne, G. (2019, December). Hierarchical motor control in mammals and machines. *Nat. Commun.*, *10*(1), 1–12.

Meyer, H. C., & Bucci, D. J. (2016, October). Neural and behavioral mechanisms of proactive and reactive inhibition. *Learn. Mem.*, *23*(10), 504–514.

Mink, J. W. (1996, November). The basal ganglia: focused selection and inhibition of competing motor programs. *Prog. Neurobiol.*, *50*(4), 381–425.

Mink, J. W., & Thach, W. T. (1991, February). Basal ganglia motor control. II. late pallidal timing relative to movement onset and inconsistent pallidal coding of movement parameters. *J. Neurophysiol.*, *65*(2), 301–329.

Mink, J. W., & Thach, W. T. (1993, December). Basal ganglia intrinsic circuits and their role in behavior. *Curr. Opin. Neurobiol.*, *3*(6), 950–957.

Miyashita, N., Hikosaka, O., & Kato, M. (1995, June). Visual hemineglect induced by unilateral striatal dopamine deficiency in monkeys. *Neuroreport*, *6*(9), 1257–1260.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... Hassabis, D. (2015, February). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533.

Mogenson, G. J., Jones, D. L., & Yim, C. Y. (1980). From motivation to action: functional interface between the limbic system and the motor system. *Prog. Neurobiol.*, *14*(2-3), 69–97.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996, March). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *J. Neurosci.*, *16*(5), 1936–1947.

Monteiro, T., Rodrigues, F. S., Pexirra, M., Cruz, B. F., Gonçalves, A. I., Rueda-Orozco, P. E., & Paton, J. J. (2020, October). *Using temperature to analyse the neural basis of a latent temporal decision.*

Moore, B. R. (2004, May). The evolution of learning. *Biol. Rev. Camb. Philos. Soc.*, *79*(2), 301–335.

Morand-Ferron, J. (2017, August). Why learn? the adaptive value of associative learning in wild populations. *Current Opinion in Behavioral Sciences*, *16*, 73–79.

Morel, A., Liu, J., Wannier, T., Jeanmonod, D., & Rouiller, E. M. (2005, February). Divergence and convergence of thalamocortical projections to premotor and supplementary motor cortex: a multiple tracing study in the macaque monkey. *Eur. J. Neurosci.*, *21*(4), 1007–1029.

Mosher, C. P., Mamelak, A. N., Malekmohammadi, M., Pouratian, N., & Rutishauser, U. (2021, January). Distinct roles of dorsal and ventral subthalamic neurons in action selection and cancellation. *Neuron*.

Motiwala, A., Soares, S., Atallah, B. V., Paton, J. J., & Machens, C. K. (2020, May). *Dopamine responses reveal efficient coding of cognitive variables.*

Mountcastle, V. B., Steinmetz, M. A., & Romoa, R. (1975). Frequency discrimination in the sense of flutter: Psychophysical measurements correlated with postcentral events in behaving monkeys. , *1969*, 1972.

Munoz, D. P., & Everling, S. (2004, March). Look away: the anti-saccade task and the voluntary control of eye movement. *Nat. Rev. Neurosci.*, *5*(3), 218–228.

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., ... Wang, X.-J. (2014, December). A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.*, *17*(12), 1661–1663.

Mussa-Ivaldi, F. A., & Bizzi, E. (2000, December). Motor learning through the combination of primitives. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *355*(1404), 1755–1769.

Nagel, G., Szellas, T., Huhn, W., Kateriya, S., Adeishvili, N., Berthold, P., ... Bamberg, E. (2003, November). Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl. Acad. Sci. U. S. A.*, *100*(24), 13940–13945.

Nakamura, K., & Hikosaka, O. (2006, December). Facilitation of saccadic eye movements by postsaccadic electrical stimulation in the primate caudate. *J. Neurosci.*, *26*(50), 12885–12895.

Nambu, A., Tokuno, H., Hamada, I., Kita, H., Imanishi, M., Akazawa, T., ... Hasegawa, N. (2000, July). Excitatory cortical inputs to pallidal neurons via the subthalamic nucleus in the monkey. *J. Neurophysiol.*, *84*(1), 289–300.

Nambu, A., Tokuno, H., & Takada, M. (2002, June). Functional significance of the cortico–subthalamo–pallidal 'hyperdirect' pathway. *Neurosci. Res.*, *43*(2), 111–117.

Nauta, H. J., & Cole, M. (1978, July). Efferent projections of the subthalamic nucleus: an autoradiographic study in monkey and cat. *J. Comp. Neurol.*, *180*(1), 1–16.

Neftci, E. O., & Averbeck, B. B. (2019, March). Reinforcement learning in artificial and biological systems. *Nature Machine Intelligence*, *1*(3), 133–143.

Nicola, S. M., & Malenka, R. C. (1997, August). Dopamine depresses excitatory and inhibitory synaptic transmission by distinct mechanisms in the nucleus accumbens. *J. Neurosci.*, *17*(15), 5697–5710.

Nicola, S. M., Surmeier, J., & Malenka, R. C. (2000). Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. *Annu. Rev. Neurosci.*, *23*, 185–215.

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychol. Bull.*, *89*(1), 133–162.

Nigg, J. T. (2001, October). Is ADHD a disinhibitory disorder? *Psychol. Bull.*, *127*(5), 571–598.

Niv, Y. (2019, October). Learning task-state representations. *Nat. Neurosci.*, *22*(10), 1544–1553.

Niv, Y. (2021, October). The primacy of behavioral research for understanding the brain. *Behav. Neurosci.*, *135*(5), 601–609.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007, April). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507–520.

Nóbrega-Pereira, S., Gelman, D., Bartolini, G., Pla, R., Pierani, A., & Marín, O. (2010, February). Origin and molecular specification of globus pallidus neurons. *J. Neurosci.*, *30*(8), 2824–2834.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004, April). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*(5669), 452–454.

Panigrahi, B., Martin, K. A., Li, Y., Graves, A. R., Vollmer, A., Olson, L., ... Dudman, J. T. (2015, September). Dopamine is required for the neural representation and control of movement vigor. *Cell*, *162*(6), 1418–1430.

Parent, A., & Hazrati, L. N. (1995, January). Functional anatomy of the basal ganglia. II. the place of subthalamic nucleus and external pallidum in basal ganglia circuitry. *Brain Res. Brain Res. Rev.*, *20*(1), 128–154.

Park, J., Coddington, L. T., & Dudman, J. T. (2020, July). Basal ganglia circuits for action specification. *Annu. Rev. Neurosci.*, *43*, 485–507.

Parker, J. G., Marshall, J. D., Ahanonu, B., Wu, Y.-W., Kim, T. H., Grewe, B. F., ... Schnitzer, M. J. (2018, May). Diametric neural ensemble dynamics in parkinsonian and dyskinetic states. *Nature*, *557*(7704), 177–182.

Passingham, R. E. (1973). Anatomical differences between the neocortex of man and other primates. *Brain Behav. Evol.*, *7*(5), 337–359.

Peak, J., Chieng, B., Hart, G., & Balleine, B. W. (2020, November). Striatal direct and indirect pathway neurons differentially control the encoding and updating of goal-directed learning. *Elife*, *9*.

Penney, J. B., Jr, & Young, A. B. (1983). Speculations on the functional anatomy of basal ganglia disorders. *Annu. Rev. Neurosci.*, *6*, 73–94.

Percheron, G., François, C., Talbi, B., Meder, J. F., Fenelon, G., & Yelnik, J. (1993). *The primate motor thalamus analysed with reference to subcortical afferent territories* (Vol. 60) (No. 1-3).

Percheron, G., François, C., Yelnik, J., Fénelon, G., & Talbi, B. (1994). The basal ganglia related system of primates: Definition, description and informational analysis. In G. Percheron, J. S. McKenzie, & J. Féger (Eds.), *The basal ganglia IV: New ideas and data on structure and function* (pp. 3–20). Boston, MA: Springer US.

Pisanello, F., Mandelbaum, G., Pisanello, M., Oldenburg, I. A., Sileo, L., Markowitz, J. E., ... Sabatini, B. L. (2017, August). Dynamic illumination of spatially restricted or large brain volumes via a single tapered optical fiber. *Nat. Neurosci.*, *20*(8), 1180–1188.

Planert, H., Szydlowski, S. N., Hjorth, J. J. J., Grillner, S., & Silberberg, G. (2010, March). Dynamics of synaptic transmission between fast-spiking interneurons and striatal projection neurons of the direct and indirect pathways. *J. Neurosci.*, *30*(9), 3499–3507.

Polyakova, Z., Chiken, S., Hatanaka, N., & Nambu, A. (2020, September). Cortical control of subthalamic neuronal activity through the hyperdirect and indirect pathways in monkeys. *J. Neurosci.*, *40*(39), 7451–7463.

Postle, B. R. (2006, April). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23–38.

Prescott, T. J., Bryson, J. J., & Seth, A. K. (2007, September). Introduction. modelling natural action selection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *362*(1485), 1521–1529.

Prescott, T. J., Montes González, F. M., Gurney, K., Humphries, M. D., & Redgrave, P. (2006, January). A robot model of the basal ganglia: behavior and intrinsic processing. *Neural Netw.*, *19*(1), 31–61.

Prescott, T. J., Redgrave, P., & Gurney, K. (1999, January). Layered control architectures in robots and vertebrates. *Adapt. Behav.*, *7*(1), 99–127.

Qiu, A., Crocetti, D., Adler, M., Mahone, E. M., Denckla, M. B., Miller, M. I., & Mostofsky, S. H. (2009, January). Basal ganglia volume and shape in children with attention deficit hyperactivity disorder. *Am. J. Psychiatry*, *166*(1), 74–82.

Rausell, E., Bickford, L., Manger, P. R., Woods, T. M., & Jones, E. G. (1998, June). Extensive divergence and convergence in the thalamocortical projection to monkey somatosensory cortex. *J. Neurosci.*, *18*(11), 4216–4232.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, *89*(4), 1009–1023.

Reig, R., & Silberberg, G. (2014, September). Multisensory integration in the mouse striatum. *Neuron*, *83*(5), 1200–1212.

Reiner, A., Albin, R. L., Anderson, K. D., D'Amato, C. J., Penney, J. B., & Young, A. B. (1988, August). Differential loss of striatal projection neurons in huntington disease. *Proc. Natl. Acad. Sci. U. S. A.*, *85*(15), 5733–5737.

Rennaker, R. L., Miller, J., Tang, H., & Wilson, D. A. (2007, June). Minocycline increases quality and longevity of chronic neural recordings. *J. Neural Eng.*, *4*(2), L1–5.

Reynolds, J. N., Hyland, B. I., & Wickens, J. R. (2001, September). A cellular mechanism of reward-related learning. *Nature*, *413*(6851), 67–70.

Reynolds, J. N. J., & Wickens, J. R. (2002, June). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw.*, *15*(4-6), 507–521.

Roesch, M. R., Calu, D. J., & Schoenbaum, G. (2007, December). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat. Neurosci.*, *10*(12), 1615–1624.

Rogers, R. D. (2011, January). The roles of dopamine and serotonin in decision making: evidence from pharmacological experiments in humans. *Neuropsychopharmacology*, *36*(1), 114–132.

Romanelli, P., Esposito, V., Schaal, D. W., & Heit, G. (2005, February). Somatotopy in the basal ganglia: experimental and clinical evidence for segregated sensorimotor channels. *Brain Res. Brain Res. Rev.*, *48*(1), 112–128.

Roos, R. A. C. (2010, December). Huntington's disease: a clinical review. *Orphanet J. Rare Dis.*, *5*, 40.

Roseberry, T. K., Lee, A. M., Lalive, A. L., Wilbrecht, L., Bonci, A., & Kreitzer, A. C. (2016, January). Cell-Type-Specific control of brainstem locomotor circuits by basal ganglia. *Cell*, *164*(3), 526–537.

Rueda-Orozco, P. E., & Robbe, D. (2015, March). The striatum multiplexes contextual and kinematic information to constrain motor habits execution. *Nat. Neurosci.*, *18*(3), 453–460.

Sahota, M. K. (1994). Action selection for robots in dynamic environments through inter-behaviour bidding. *From animals to animats*, *3*, 138–142.

Sapp, E., Ge, P., Aizawa, H., Bird, E., Penney, J., Young, A. B., ... DiFiglia, M. (1995, January). Evidence for a preferential loss of enkephalin immunoreactivity in the external globus pallidus in low grade huntington's disease using high resolution image analysis. *Neuroscience*, *64*(2), 397–404.

Sato, F., Lavallée, P., Lévesque, M., & Parent, A. (2000, January). Single-axon tracing study of neurons of the external segment of the globus pallidus in primate. *J. Comp. Neurol.*, *417*(1), 17–31.

Saxena, S., & Rauch, S. L. (2000, September). Functional neuroimaging and the neuroanatomy of obsessive-compulsive disorder. *Psychiatr. Clin. North Am.*, *23*(3), 563–586.

Schevernels, H., Bombeke, K., Van der Borght, L., Hopf, J.-M., Krebs, R. M., & Boehler, C. N. (2015, November). Electrophysiological evidence for the involvement of proactive and reactive control in a rewarded stop-signal task. *Neuroimage*, *121*, 115–125.

Schmidt, R., Leventhal, D. K., Mallet, N., Chen, F., & Berke, J. D. (2013, August). Canceling actions involves a race between basal ganglia pathways. *Nat. Neurosci.*, *16*(8), 1118–1124.

Schroll, H., & Hamker, F. H. (2013, December). Computational models of basal-ganglia pathway functions: focus on functional neuroanatomy. *Front. Syst. Neurosci.*, *7*, 122.

Schultz, W. (1995). *The primate basal ganglia between the intention and outcome of action.*

Schultz, W., Apicella, P., & Ljungberg, T. (1993, March). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.*, *13*(3), 900–913.

Schultz, W., Dayan, P., & Montague, P. R. (1997, March). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Schwarting, R. K., & Huston, J. P. (1996, October). The unilateral 6-hydroxydopamine lesion model in behavioral brain research. analysis of functional deficits, recovery and treatments. *Prog. Neurobiol.*, *50*(2-3), 275–331.

Searle, S. R., Speed, F. M., & Milliken, G. A. (1980, November). Population marginal means in the linear model: An alternative to least squares means. *Am. Stat.*, *34*(4), 216–221.

Selemon, L. D., & Goldman-Rakic, P. S. (1985, March). Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. *J. Neurosci.*, *5*(3), 776–794.

Shen, W., Flajolet, M., Greengard, P., & James Surmeier, D. (2008). *Dichotomous dopaminergic control of striatal synaptic plasticity* (Tech. Rep.).

Shiflett, M. W., Brown, R. A., & Balleine, B. W. (2010, February). Acquisition and performance of goal-directed instrumental actions depends on ERK signaling in distinct regions of dorsal striatum in rats. *J. Neurosci.*, *30*(8), 2951–2959.

Siegle, J. H., López, A. C., Patel, Y. A., Abramov, K., Ohayon, S., & Voigts, J. (2017, August). Open ephys: an open-source, plugin-based platform for multichannel electrophysiology. *J. Neural Eng.*, *14*(4), 045003.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.*, *24*, 1193–1216.

Simonyan, K. (2019, January). Recent advances in understanding the role of the basal ganglia. *F1000Res.*, *8*.

Sippy, T., Lapray, D., Crochet, S., & Petersen, C. C. H. (2015, October). Cell-Type-Specific sensorimotor processing in striatal projection neurons during Goal-Directed behavior. *Neuron*, *88*(2), 298–305.

Sjöbom, J., Tamtè, M., Halje, P., Brys, I., & Petersson, P. (2020, October). Cortical and striatal circuits together encode transitions in natural behavior. *Sci Adv*, *6*(41).

Skinner, B. F. (1938). The behavior of organisms: an experimental analysis. , *457*.

Smith, J. M. (1979, September). Game theory and the evolution of behaviour. *Proc. R. Soc. Lond. B Biol. Sci.*, *205*(1161), 475–488.

Smith, Y., Bevan, M. D., Shink, E., & Bolam, J. P. (1998, September). Microcircuitry of the direct and indirect pathways of the basal ganglia. *Neuroscience*, *86*(2), 353–387.

Soares, S., Atallah, B. V., & Paton, J. J. (2016, December). Midbrain dopamine neurons control judgment of time. *Science*, *354*(6317), 1273–1277.

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017, November). The hippocampus as a predictive map. *Nat. Neurosci.*, *20*(11), 1643–1653.

Starr, P. A., Kang, G. A., Heath, S., Shimamoto, S., & Turner, R. S. (2008, May). Pallidal neuronal discharge in huntington's disease: support for selective loss of striatal cells originating the indirect pathway. *Exp. Neurol.*, *211*(1), 227–233.

Starr, P. A., Rau, G. M., Davis, V., Marks, W. J., Jr, Ostrem, J. L., Simmons, D., ... Turner, R. S. (2005, June). Spontaneous pallidal neuronal activity in human dystonia: comparison with parkinson's disease and normal macaque. *J. Neurophysiol.*, *93*(6), 3165–3176.

Steiner, H., & Tseng, K. Y. (2016). *Handbook of basal ganglia structure and function*. Academic Press.

Steinmetz, N. A., Zatka-Haas, P., Carandini, M., & Harris, K. D. (2019, December). Distributed coding of choice, action and engagement across the mouse brain. *Nature*, *576*(7786), 266–273.

Stephenson-Jones, M., Samuelsson, E., Ericsson, J., Robertson, B., & Grillner, S. (2011, July). Evolutionary conservation of the basal ganglia as a common vertebrate mechanism for action selection. *Curr. Biol.*, *21*(13), 1081–1091.

Sterling, P., & Laughlin, S. (2017). *Principles of neural design*. MIT Press.

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., & Harris, K. D. (2019, July). High-dimensional geometry of population responses in visual cortex. *Nature*, *571*(7765), 361–365.

Sutton, R. S., Barto, A. G., & Others. (1998). *Introduction to reinforcement learning* (Vol. 135). MIT press Cambridge.

Sutton, R. S., Precup, D., & Singh, S. (1999, August). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artif. Intell.*, *112*(1), 181–211.

Tai, L.-H., Lee, A. M., Benavidez, N., Bonci, A., & Wilbrecht, L. (2012, September). Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat. Neurosci.*, *15*(9), 1281–1289.

Taverna, S., Ilijic, E., & Surmeier, D. J. (2008, May). Recurrent collateral connections of striatal medium spiny neurons are disrupted in models of parkinson's disease. *J. Neurosci.*, *28*(21), 5504–5512.

Tecuapetla, F., Jin, X., Lima, S. Q., & Costa, R. M. (2016, July). Complementary contributions of striatal projection pathways to action initiation and execution. *Cell*, *166*(3), 703–715.

Tecuapetla, F., Matias, S., Dugue, G. P., Mainen, Z. F., & Costa, R. M. (2014, July). Balanced activity in basal ganglia projection pathways is critical for contraversive movements. *Nat. Commun.*, *5*, 4315.

Tepper, J. M., & Bolam, J. P. (2004, December). Functional diversity and specificity of neostriatal interneurons. *Curr. Opin. Neurobiol.*, *14*(6), 685–692.

Thorn, C. A., Atallah, H., Howe, M., & Graybiel, A. M. (2010, June). Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron*, *66*(5), 781–795.

Thorndike, E. L. (1911). *Animal intelligence; experimental studies, by edward l. thorndike.*

Tierney, A. J. (1986, December). The evolution of learned and innate behavior: Contributions from genetics and neurobiology to a theory of behavioral evolution. *Anim. Learn. Behav.*, *14*(4), 339–348.

Tobler, P. N., O'Doherty, J. P., Dolan, R. J., & Schultz, W. (2007, February). Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J. Neurophysiol.*, *97*(2), 1621–1632.

Todorov, E., Li, W., & Pan, X. (2005, November). From task parameters to motor synergies: A hierarchical framework for approximately-optimal control of redundant manipulators. *J. Robot. Syst.*, *22*(11), 691–710.

Turner, R. S., & Desmurget, M. (2010, December). Basal ganglia contributions to motor control: a vigorous tutor. *Curr. Opin. Neurobiol.*, *20*(6), 704–716.

van der Kooy, D., & Carter, D. A. (1981, April). The organization of the efferent projections and striatal afferents of the entopeduncular nucleus and adjacent areas in the rat. *Brain Res.*, *211*(1), 15–36.

van der Meer, M. A. A., Johnson, A., Schmitzer-Torbert, N. C., & Redish, A. D. (2010, July). Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. *Neuron*, *67*(1), 25–32.

van Wouwe, N. C., Neimat, J. S., van den Wildenberg, W. P. M., Hughes, S. B., Lopez, A. M., Phibbs, F. T., ... Wylie, S. A. (2020, November). Subthalamic nucleus subregion stimulation modulates inhibitory control. *Cereb Cortex Commun*, *1*(1), tgaa083.

Vitek, J. L. (2002). Pathophysiology of dystonia: a neuronal model. *Mov. Disord.*, *17 Suppl 3*, S49–62.

von Uexküll, J. (2013). *A foray into the worlds of animals and humans: with a theory of meaning*. U of Minnesota Press.

Voon, V., Gao, J., Brezing, C., Symmonds, M., Ekanayake, V., Fernandez, H., ... Hallett, M. (2011, May). Dopamine agonists and risk: impulse control disorders in parkinson's disease. *Brain*, *134*(Pt 5), 1438–1446.

Wagner, G. P., & Altenberg, L. (1996, June). PERSPECTIVE: COMPLEX ADAPTATIONS AND THE EVOLUTION OF EVOLVABILITY. *Evolution*, *50*(3), 967–976.

Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., ... Botvinick, M. (2018, June). Prefrontal cortex as a meta-reinforcement learning system. *Nat. Neurosci.*, *21*(6), 860–868.

Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A., & Uchida, N. (2012, June). Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron*, *74*(5), 858–873.

Watanabe, M., & Munoz, D. P. (2010, July). Presetting basal ganglia for volitional actions. *J. Neurosci.*, *30*(30), 10144–10157.

Wichmann, T., Bergman, H., Starr, P. A., Subramanian, T., Watts, R. L., & DeLong, M. R. (1999, April). Comparison of MPTP-induced changes in spontaneous neuronal discharge in the internal pallidal segment and in the substantia nigra pars reticulata in primates. *Exp. Brain Res.*, *125*(4), 397–409.

Willis, T. (1685). *The london practice of physick: Or the whole practical part of physick ... faithfully made into english and printed for the public good*. Printed at the George in Fleet Street.

Yael, D., Zeef, D. H., Sand, D., Moran, A., Katz, D. B., Cohen, D., ... Bar-Gad, I. (2013, December). Haloperidol-induced changes in neuronal activity in the striatum of the freely moving rat. *Front. Syst. Neurosci.*, *7*, 110.

Yagishita, S., Hayashi-Takagi, A., Ellis-Davies, G. C. R., Urakubo, H., Ishii, S., & Kasai, H. (2014, September). A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science*, *345*(6204), 1616–1620.

Yeterian, E. H., & Pandya, D. N. (1991, October). Prefrontostriatal connections in relation to cortical architectonic organization in rhesus monkeys. *J. Comp. Neurol.*, *312*(1), 43–67.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004, January). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.*, *19*(1), 181–189.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2006, January). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behav. Brain Res.*, *166*(2), 189–196.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005, July). The role of the dorsomedial striatum in instrumental conditioning. *Eur. J. Neurosci.*, *22*(2), 513–523.

Yoshida, A., & Tanaka, M. (2016, March). Two types of neurons in the primate globus pallidus external segment play distinct roles in antisaccade generation. *Cereb. Cortex*, *26*(3), 1187–1199.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., & Lipson, H. (2015, June). Understanding neural networks through deep visualization.

Yttri, E. A., & Dudman, J. T. (2016, May). Opponent and bidirectional control of movement velocity in the basal ganglia. *Nature*, *533*(7603), 402–406.

Zador, A. M. (2019, August). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat. Commun.*, *10*(1), 3770.

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt