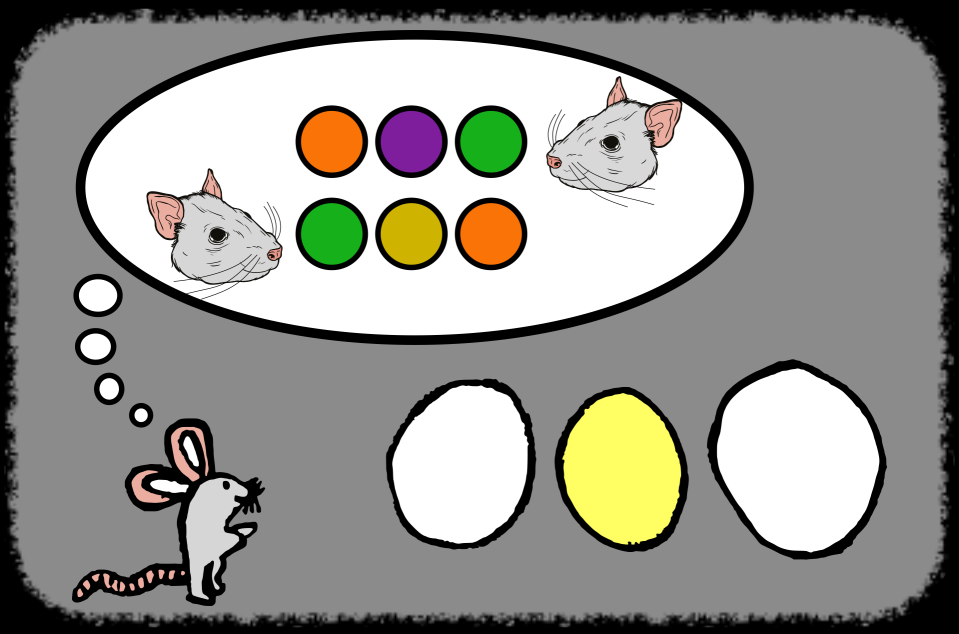
2024

Context dependent expectation signals in behavior and brain

Mauricio Alejandro Toro Espejo



Dissertation presented to obtain the **Ph.D degree in Neuroscience**

International Neuroscience Doctoral Programme

Oeiras, February, 2024

INSTITUTO DE TECNOLOGIA QUÍMICA E BIOLÓGICA ANTÓNIO XAVIER | UNIVERSIDADE NOVA DE LISBOA

[This Page Intentionally Not Left Blank]

Context dependent expectation signals in behavior and brain

Mauricio Alejandro Toro Espejo

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

Oeiras, February, 2024

Context dependent expectation signals in behavior and brain

Mauricio Alejandro Toro Espejo

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

Supervised by: Joseph J. Paton
International Neuroscience Doctoral Programme
Champalimaud Research
Lisbon, Portugal

2024

Cover artwork by Ramona and Mauricio Toro, 2022.
Vectorized rat face is adapted from an original CC-BY image from StockUnlimited, image ID: 1425973.

[This Page Intentionally Left Not Blank]

List of Abbreviations

AUC-ROC Area under the receiver operating characteristics

BG basal Ganglia

CBGTC Cortico-Basal Ganglia-Thalamo-Cortical

DA Dopamine

EP Entopeduncular nucleus

GPi External segment of the Globus Pallidus

GPi Internal segment of the Globus Pallidus

ITOI Inter trial onset interval

MTh Motor Thalamus

MUA Multi unit activity

PETH Peri-event time histogram

RL Reinforcement learning

SNc Substantia Nigra pars Compacta

SNr Substantia Nigra pars Reticulata

SU Single unit

SVM Support vector machine

Thal Thalamus

VA/VL Ventral-Anterior & -Lateral nuclei of the Thalamus

VTA Ventral Tegmental Area

Abstract

We recognize that the world has structure, and take advantage of that to be prepared to act given what we have done previously in similar situations and the outcomes of those actions. By interacting with our world, we learn that in different contexts some actions are more adaptive than others. Thus, once we recognize the situation that we are in at a given time, we can have some action prepared, a default plan. The cortico-basal ganglia-thalamo-cortical loop (CBGTC) is thought to be in charge of the recognition of context, and controlling the expression of behaviors depending on their expected outcomes, v.g. the amount of vigor to exert in the execution; and dysregulation of their function are associated with neurological and psychiatric disorders. Critically, given the circuit connectivity, the output of the basal ganglia (BG), the *subtantia nigra pars reticulata* (SNr), is the most direct way that BG reinforcement-learning like algorithm can use to communicate with cortex, via this region inhibitory projections to the motor thalamus (MTh). However, little is known about the population level computations of these regions in the process of preparation of orienting movements for different reinforcers.

We developed a behavioral assay that enables rats to demonstrate their comprehension of their environment by adjusting the intensity of their actions. This assay also requires them to update their default strategies and enhance their understanding of the context. In a cohort of animals, we monitored signals from the primary output of the basal ganglia (BG) and the motor thalamus (MTh) while they engaged in this task. Our observations revealed signals related to their understanding of various task dimensions during key behavioral events. Additionally, we observed differences in the stability of these signals, indicating specific roles for these regions during relevant periods, aligning with their anatomical locations and projection patterns within the CBGTC.

Our findings contribute significantly to our understanding of the roles of the SNr and the MTh in behavioral control and motor preparation. Further exploration of our task and results promises a deeper and broader comprehension of the CBGTC, which could eventually lead to targeted interventions benefiting patients with neurological or psychiatric conditions linked to dysregulation of the activity in these brain regions.

Sinais de expectativa dependentes do contexto no comportamento e no cérebro

Reconhecemos que o mundo tem estrutura e aproveitamos isso para estarmos preparados para agir, tendo em conta o que fizemos anteriormente em situações semelhantes e os resultados dessas ações. Ao interagir com o nosso mundo, aprendemos que em diferentes contextos algumas ações são mais adaptativas do que outras. Assim, uma vez que reconhecemos a situação em que nos encontramos num determinado momento, podemos ter alguma ação preparada, uma estratégia padrão. Acredita-se que o loop córtico-gânglios basais-tálamo-cortical (CBGTC) seja responsável pelo reconhecimento do contexto e pelo controle da expressão de comportamentos dependendo de seus resultados esperados, v.g. a quantidade de vigor a exercer na execução; e a desregulação da sua função estão associadas a distúrbios neurológicos e psiquiátricos. Criticamente, dada a conectividade do circuito, a saída dos gânglios da base (BG), a *substantia nigra pars reticulata* (SNr), é a maneira mais direta que o algoritmo de aprendizagem por reforço BG pode usar para se comunicar com o córtex, através desta região, projeções inibitórias para o tálamo motor (MTh). Porém, pouco se sabe sobre os cálculos do nível populacional dessas regiões no processo de preparação de movimentos de orientação para diferentes reforçadores.

Desenvolvemos um ensaio comportamental que permite aos ratos demonstrar a sua compreensão do seu ambiente, ajustando a intensidade das suas ações. Este ensaio também exige que atualizem as suas estratégias padrão e melhorem a sua compreensão do contexto. Em uma coorte de animais, monitoramos sinais da saída primária dos gânglios da base (BG) e do tálamo motor (MTh) enquanto eles realizavam essa tarefa. Nossas observações revelaram sinais relacionados à compreensão de várias dimensões da tarefa durante eventos comportamentais importantes. Além disso, observamos diferenças na estabilidade desses sinais, indicando papéis específicos para essas regiões durante períodos relevantes, alinhando-se com suas localizações anatômicas e padrões de projeção dentro da alça córtico-basal gânglios-tálamo-cortical CBGTC

Os nossos resultados contribuem significativamente para a nossa compreensão dos papéis da SNr e do MTh no controle comportamental e na preparação motora. Uma exploração mais aprofundada da nossa tarefa e dos resultados promete uma compreensão mais profunda e mais ampla do CBGTC, o que poderia eventualmente levar a intervenções direcionadas que beneficiam pacientes com condições neurológicas ou psiquiátricas ligadas à desregulação da atividade nestas regiões cerebrais.

Author Contributions

All chapter in the present thesis were written by Mauricio Toro. Experiments in chapter 2 and 3 were designed and implemented by Mauricio Toro, Filipe Rodrigues, Tiago Monteiro, and Joseph Paton. Experimental data was aquired by Mauricio Toro. Sofia Castro e Almeida, Tiago Monteiro, Filipe Rodrigues, Margarida Pexirra, Sofia Freitas and Renato Sousa assisted with the training and collection of behavioral and/or electrophysiological recording sessions included in the datasets. Data processing and analysis was carried out by Mauricio Toro with supervision and assistance from Joseph Paton. The linear accumulator to threshold model presented in chapter 4 was design by Mauricio Toro, Filipe Rodrigues, Tiago Monteiro, Margarida Sousa, and Sofia Castro e Almeida, with the supervision and assistance of Joseph Paton. The asynchronous-advantage-actor-critic model and environment presented in chapter 4 were developed and implemented by Mauricio Toro and Margarida Sousa, with the supervision and assistance of Joseph Paton.

Financial Support

Funding for this work was provided by the Portuguese FCT (Fundação para a Ciência e a Tecnologia, FCT Bolsa PD/BD/114275/2016) and Fundação Champalimaud. The work was carried out within the International Neuroscience Doctoral Program (INDP) 2015.

Contents

Chapter 1

Introduction

He said: "It's all in your head"; and
I said: "So's everything", but he
didn't get it.

Paperbag,
Fiona Apple

1.1 General introduction

We say that we are in a particular context when we understand that the current state of the world shares similarities with others that we have experienced, and these imply something about what we ought to do. Understanding that similar configurations of the world can imply that different sets of actions can be more beneficial, allows us to be prepared to respond in the most advantageous manner. In a world where we are driven to act, having a way to bias some actions given our previous experiences allows us to get the most of novel situations. Taking this into account, our working definition of context will be the set of underlying circumstances that gives meaning to particular action-outcomes mappings. In this sense, for the subjects in our task, where there are discrete actions, in the sense of motor target goals, in each context, only one action will be the most rewarding. We set sails to find if animals learn to use this information to guide their behavior, and what are the neural correlates of this in two key regions in the circuitry relevant for the processing and communication of these signals.

During this introductory chapter, we define the core ideas that will appear throughout the dissertation. In section 1.2 we describe how animals engage with

environment through actions, with focus on behavioral assays and our interpretation of their roles in the problem. Section 1.3 introduces the brain mechanism associated to behavioral control, in general the cortico-basal ganglia-thalamo-cortical (CBGTC) loop, and in particular the roles of basal ganglia (BG) and motor thalamus (MTh). Later, in section 1.4 we introduce context-guided behavioral control, and describe the task that inspires the present work. Finally, in section 1.5 we conclude by stating the objectives of our research, presenting the question that guide the design of our study, and the laid down the set of hypothesis to be tested in the following chapters.

1.2 Acting in an environment

To interact with the world, animals are endowed with an abundance of sensory systems to receive and integrate perturbations in their external and internal environments, making meaningful-for-themselves relations that allow them to use the effectors which their evolutionary niche has endowed them with to pursue their objectives (von Uexküll and von Uexküll, 2010). Their objectives can be as simple as survival and maintenance of homeostatic equilibrium, or complicated as completing the necessary work to present and defend a doctoral dissertation or writing a scientific article; as they can carry significant barriers to be surpassed, and the latter has proven that some barriers can be found across many cultural backgrounds (Upper, 1974; Didden et al., 2007). One way to think about objectives, is to consider them as the teleological cause for actions, where the expected outcome associated to them, the reinforcer, calls for their approach or avoidance, by their value to the organism (Aarts and Elliot, 2012, Chap. 1). The *call* they make is what will drive the animals to them. These calls for actions can be responded to in different ways, either in an automatic and unreflective manner or with crafted and organized control. These distinctive types of action control are called habitual or goal-directed, respectively. On the one hand, habitual responses are simple, almost stereotypical behaviors, which after an external or internal signal are executed swiftly. On the other, goal-directed behaviors, require to consider the environment and situation *at hand*, planning a suitable response and their execution. In well-know situations, animals resort to habitual behaviors, as these require a lower metabolic cost. But, if a new situation or a large enough perturbation arises, they will evaluate other means to regain a sense of coupling with their environment[1]. Taking into account their knowledge about the environment, the available actions and how they relate to their goals, i.e., by performing goal-directed actions. But, before being able to take actions, an agent needs to be aware of their environment, getting to know

the physical surroundings and the available and adaptive actions that are in their disposal.

1.2.1 Learning about what is out there

If we think of a naïve and satiated animal, presented in a new, non-threatening environment. At first, the animals will explore their surroundings until he has acknowledged and gained enough information about it to stay minding his own animal business. An example of this kind of coupling is the novel-object recognition task (presented in Ennaceur and Delacour (1988), depicted in Fig. 1.1). This is a one-shot learning task, where a rat is situated in a box with a number of sample objects for a period of time; after this, they are removed from the enclosure, and one of the objects is changed by a new one; finally, the animal is again moved inside the enclosure. In this task, the variable to be read from behavior is the time the animal spend interacting or exploring the different objects. In the first exposure, animals explore all objects about the same amount of time, but after one of them is changed, they are more likely to explore the novel one far longer than those that did not vary. This type of task has been used to explore the development of recognition memory in rats (Reger et al., 2009), and study the roles of different brain regions and neuromodulators in these process (Orofino et al., 1999; Alvarez and a Alvarez, 2008; Osorio-Gómez et al., 2022; Okada et al., 2022). We are driven by our interest in understanding the relationship between animals and their objectives, and how these goals drive their behavioral interaction with environments. But, in this task there is no primary reinforcer, no need to be satisfied, nor danger to be avoided. Nonetheless, animals have a propensity for exploration, and are generally curious (Berlyne, 1955), and acquiring information is valuable for them (Ajuwon et al., 2022). During the first interaction with the environment, animals will explore to gain some insights about it; after they have acknowledged what the state of this environment is, they will not need to spend more time conflicting about what is affordable to be done in this place. In the second exposure, when presented with a perturbed known environment, they will first explore the novel element, to gain insight about it. But, as the previously known does not inform, nor give anything new, it can be disregarded. At first, there is an exploration to understand the environment; later, as part of it has changed, they explore again, but the unchanged elements are non-relevant, they give no new information, the "what

[1]The concept of environmental coupling is to be thought in the sense of Maturana and Varela's *structural coupling* (Maturana and Varela, 1980), they present the idea that in the constant interaction between organism and environment, environmental perturbations present a challenge for the former, to which she has to adapt to secure her conservation.

can be done with them?" is solved. In this simple case, we can pose that the animal has first made an understanding of their environment, formed a representation of it. And, after an update on this environment, the animal explores it again to receive new information and becomes knowledgeable about it. There is value in being aware of the surroundings.

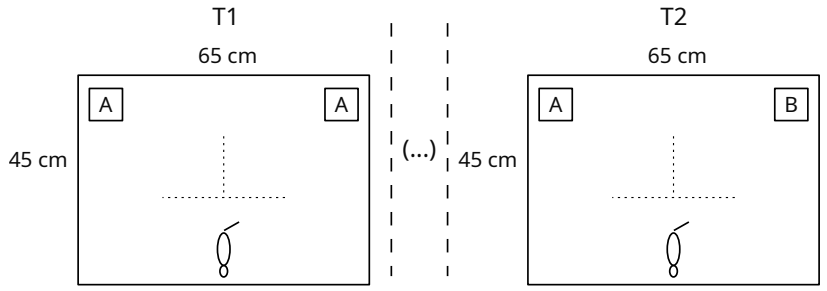


Figure 1.1: Graphical depiction of the novel-object recognition task. Diagram depicting the relevant manipulation for the task, where at first an animal is situated in an enclosure with two similar objects, T1 left panel. After a delay, one of the objects is changed, T2 right panel. The difference in time spent exploring the novel object is used as a proxy for the novelty the animal assigns to this new element. Adapted from Ennaceur and Delacour (1988).

If we were to expose naïve rats to a compromised external world, they will look for ways to solve the matter at hand. And, if the perturbation is consistently solvable in a particular manner, they will resort to this solution in new exposures. An example of external world perturbations that induces discomfort and has a consistent solution for the rat, is the Morris water maze (Morris, 2008; Nunez, 2008; Othman et al., 2022). In this task, animals are located in a circular pool with opaque water covering a submerged platform. After being introduced in the pool, the rat will swim around randomly, and eventually find the location of the submerged platform. After some number of repetitions, for any new entry location, the animal will quickly go towards the platform. This type of assay and other similar types of mazes, have proven tremendously helpful to study a broad range of behavioral phenomena and their neural correlates, from spatial navigation and memory formation (Brandeis et al., 1989; Nakagawa et al., 1995; de Bruin et al., 1997; Callaway et al., 2012). These assay induces discomfort, as the animal is required to keep swimming to be safe, once he has reached the submerged platform, he can regain a sense of safety and evaluate how this was achieved. By doing so, in case of a new encounter with this situation, he will be able to solve it quickly. Thus, their behavior is at first exploratory and later more stereotypical, moving from goal-directed to habitual; first being exploratory, and after a solution is found, requiring the use of external or internal cues to reach the plat-

form, but after enough repetitions, the "how to get there?" problem is easily solved. Animals can also learn how to use their environment to their advantage.

1.2.2 Learning what to do

We can continue now with examples where animals are in a perturbed internal world, in an environment that has elements that would allow solving these issues. In this case, animals will first engage in explorations of their surroundings, and the available actions. After finding a solution that reaches the objective, or achieves re-coupling of the agent and their environment, in future exposures they will spring to resort to this same solution. One collection of early experimental evidence of these kinds of situations, are the experiments of Edward Thorndike with animals inside wooden boxes (Thorndike (1898); Chance (1999), Fig. 1.2, left panel). In them, food-deprived cats are situated in a wood box inside a room from which they can not escape. One of the box walls has a guillotine door that can be opened from the inside if a particular action is performed. On the outside of the box lays a plate of food, enticing the food-deprived feline. During the first exposure, the cat will notice that there is no easy way out, and explore different calls and movements within the enclosure. Eventually, the cat will inadvertently perform the appropriate action, releasing himself from the box, and allowing them to reach the food, to later be put inside the enclave once again. Over repeated exposures, the time that animals take to leave the box shortens. Also, the complexity of the actions required for them to be released correlates with the time it takes them to solve the problem, and how many iterations it takes them to reliably solve it quickly. Coming back to the distinction between habitual and goal-directed behaviors, we can interpret that the cats in this example, after finding by chance the solution, are resorting to their understanding of their environment rules to solve the task; some might say that they are evaluating hypothesis about what happened and testing them. Later, as they have become used to the relation, they just produce the required behavior when needed. They have understood that in this situation, this singular action is sufficient and necessary to achieve their objective. In short, some environments require particular actions to be taken to achieve a goal.

The development of behavioral manipulations and assays by B.F. Skinner led to a new wave of studies in instrumental learning. The operant conditioning chamber, also called *Skinner box* (Fig. 1.2 right panel), allowed training animals to make new seemingly arbitrary mappings between stimuli, actions, and outcomes. A controllable box with a lever, a speaker, and some form of reinforcer dispenser, could be used to

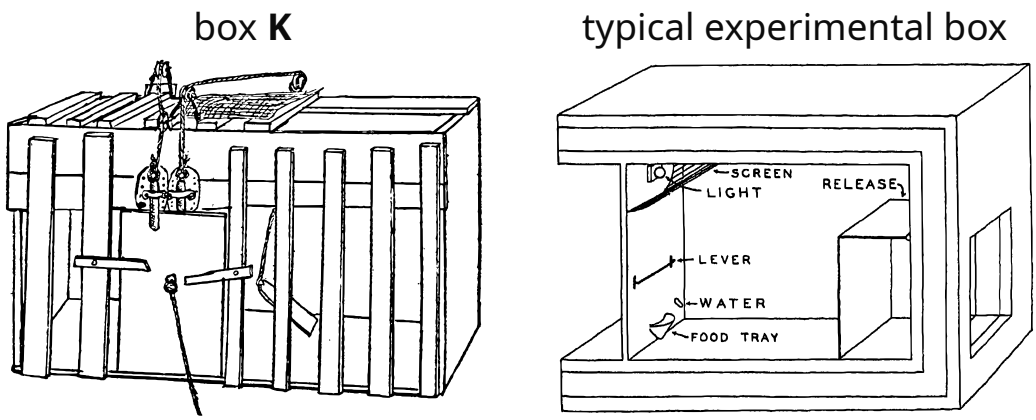


Figure 1.2: Diagrams of apparatuses pioneering the study of animal learning. The figure shows the vectorized versions of the original diagrams given by E. Thorndike (Thorndike, 1898, pg. 8) and B.F. Skinner (Skinner, 1938, pg. 49), left and right panels respectively. These kinds of apparatuses represent a foundation for the field of experimental psychology and animal learning, as they allowed the development of assays that could impose arbitrary mappings between behavior and outcomes.

investigate how adaptive behaviors are learned (Skinner, 1938). Also, these assays have work as proxy to study the development of maladaptive behavioral traits, such as substances or behavioral addictions (Thompson, 1968; Foddy, 2016; Bouton, 2021). Key insights gained from this new methods, include, amongst others, the relevancy of the reinforcement schedule habits development, and how to chain groups of behavior into a more complex one (Staddon and Cerutti, 2003).

1.2.3 Being prepared

In these examples, we have only focused on the production and expression of the behaviors. As noted earlier, a relevant part of how context can affect behavioral control is by means of anticipation, biasing or preparing an action given prior experience in the situation. These anticipatory or preparatory processes have been considered one of the most relevant features in behavioral control (Butz et al., 2008; Frese and Sabini, 2021). Brain activity signatures of preparation before acting are known to exist in intentional, self-paced human actions (Kornhuber and Deecke, 1965; Libet et al., 1983). Before the execution of an action, a number of regions in the brain are engaged up to a second before the action is taken, and their activity allows to accurately predict the timing of the action release (Schurger et al., 2021). If actions are not self-paced, but are required to be executed after the presentation of an imperative stimulus (or go-cue), we gain access to the psycho-physiological processes underlying this preparatory process. With these type of assays it has been shown how a warning

signals presented before the imperative stimulus, can facilitate the response, observed as decreases in reaction times (RT) (Bertelson, 1967; Nickerson et al., 1969; Hackley, 2009). The underlying process of this facilitation seems to be more related to response-preparation activity, than to a sensory or perceptual process (Henderson and Dittrich, 1998; Rudell and Hu, 2001). This response preparation seems to be affected by aging (Hardwick et al., 2022), and rewards are known to affect the readiness with which action would be taken (Milstein and Dorris, 2007; Shadmehr et al., 2019). One way in which rewards could impact these preparatory process is by means of the premature engagement of relevant regions needed to the execution of the action. For example, increasing corticospinal excitability (Klein et al., 2012; Bundt et al., 2016), and circuit levels cortical activity (Iigaya et al., 2020). Rewards can also facilitate being prepared by modifying the vigor to be exerted in an action. An everyday example is the time we take to answer our mobile phones when we see a call from a loved one or an unknown number. In the first case, we are eager to answer, to quickly get the rewarding sound of their voices; whereas, in the second, we might ponder if it makes sense to answer, or if it might just be another non-requested offer from some automated service. The eagerness, or fast RT for the motivating call, requires the expenditure of more energy, being more expensive for our body, we need to limit this kind of response for actions that we expect to be rewarding. In behavioral psychology, this speeding or eagerness is an effect of vigor, a dimensionless quantity that defines the amount of effort we are willing to impose in actions (Shadmehr, 2020). Actions that are expected to be more rewarding are executed quicker (Steverson et al., 2019; Choi et al., 2014), striatal dopamine levels are associated with the effort made in movements (Panigrahi et al., 2015), and neurons in the nucleus accumbens correlate with movement parameters associated to vigor (Levcik et al., 2021). For the aforementioned reasons, these motivating signals and their underlying causes, could allow animals to be prepared, introducing purpose in their plans, that could facilitate the selection of some actions. In sum, these anticipatory processes are a way of linking desirable states with a motor plan, allowing to prepare an action given their expected consequences.

As a digest of what we have described up to this point: animals act in their environments driven by their internal goals. During their interactions with their surroundings, animals learn to take adaptive actions, and in known and stable environments, they are able to prepare actions that are expected to yield better outcomes.

To engage with the environment, thus perceive, act, and learn about how actions relate to outcomes, evolution have endowed vertebrates with the complex and dynamic system inside our head—the brain. The main circuit associated with the processes of perception, action, and learning about outcomes is the cortical-basal ganglia-thalamo-

cortical (CBGTC) loop. A collection of regions that have been shown to be relevant for the acknowledgment, maintenance, and update of state representation and action-outcomes mappings (Wang et al., 2021). In the next section, we focus on the different components of this process and relate them to behavior preparation.

1.3 Neural basis of behavioral control

The cortical-basal ganglia-thalamo-cortical (CBGTC), is considered the relevant circuit underlying the recognition of states of the world, and biasing the selection of motor plans given previous experiences. The different levels of the circuit present similar organizing principles that are conserved across the involved regions. In this section we will give an overview of how these regions are relevant to guide context dependent actions, with especial consideration to the basal ganglia (BG), and motor thalamus (MTh); then, we look at relevant architectural constrains present in the CBGTC loop and what they imply for the computations that they undergo. Finally, we focus on the key nodes of interest within the circuit, as their location in the pipeline makes them promising candidates to understand how information and computations about context and outcomes expectations could be propagated from BG into cortex.

1.3.1 Structures in a loop

The cortex, BG, and thalamus are structures that form an interconnected circuit in the brain, a—mostly—closed loop. This loop plays a critical role in action preparation, execution, and learning about adaptive behaviors. In general, the cortex is responsible for integrating sensory information and higher-order processing, including decision-making and planning. BG has generally been considered to act as a gatekeeper, selecting and filtering the most appropriate action to execute, given previous experiences. The thalamus has largely being though to serve as a relay station, relaying information between the cortex and BG. Through this loop, the brain can learn from experiences, adjust behavior based on feedback, and develop adaptive strategies for achieving goals. Dysfunction in this loop has been implicated in a range of neurological and psychiatric disorders, including Parkinson's disease, addiction, and obsessive-compulsive disorder. In this section, we are going to briefly describe the roles of cortex, BG, and thalamus (figure 1.3 A), and how their interconnections facilitates the internalization of environmental state, and using previous experiences to decide action plans. Originally, the circuit was associated with motor control, given the connectivity of the structure with motor controlling territories and historical evidence

of pathologies of movement control (Fig. 1.3 B) (Alexander et al., 1986; Mink and Thach, 1993). But, the intrinsic dopaminergic (DA) nuclei within BG also encodes relevant signals related to value, reward expectation or salience (Wise and Bozarth, 1982; Schultz et al., 1993; Wise, 2009; Berridge, 2007). Including these signals in the circuit made it richer (Fig. 1.3 C), as they allowed the selection, biasing, or gating of actions given previously experienced results, and aggregates reward expectation as a factor to these processes (Graybiel et al., 1994; Kimura, 1995; Graybiel, 1995). There is a conserved hierarchical structure within the regions, with their connectivity patterns, woven by a myriad of neuromodulators in the present-day synthesis of the CBGTC (fig. 1.3 D). This makes this circuit a foremost candidate to be the central nervous system network in charge of setting, maintaining and updating agent-centric state representations, that allows to prepare and execute actions given previously experienced results (Foster et al., 2021; Wei and Wang, 2016; Sych et al., 2022). The diagrams make clear that the shortest path that BG computations have to reach cortex are the substantia nigra pars reticulata (SNr) and the internal segment of the globus pallidus (GPi) projections into thalamic nuclei.

1.3.2 Brief anatomy of cortical territories

The perception of our environment requires the interactions of many level of our central nervous system, and different cortical regions have been considered by large the most relevant ones to sustain our awareness of them (Bennett and Hacker, 2012). The external mantle of our cerebrum in his many folds and wrinkles is organized in functional territories, where mainly three categories are generally considered: primary sensory, higher-order sensory and associative regions (Kandel, 2013, Chap. V). In the sensory regions, thalamic input from the sensory periphery is processed and integrated hierarchically, where primary regions receive small portions of the stimulus and higher order region are able to relate them into more abstract and general elements. At each level of the cortical hierarchy, sensory information is organized in a parallel and structured manner. For example, the lateral geniculate nucleus of the thalamus, visual thalamus, projects into the primary visual cortex, V1. And these projections are organized in such a manner that neighboring receptive cells in the retina project to neighboring cortical neurons, generating a retinotopic map of the receptive field in V1 (Fig. 1.4 A). This structured organization is also present in the body map in mammalian motor and somatosensory cortices (Fig. 1.4 B). This architecture implies that, within cortical territories, there are relevant mappings for external/internal environmental elements running in parallel, at different granularity. Whilst, abstraction

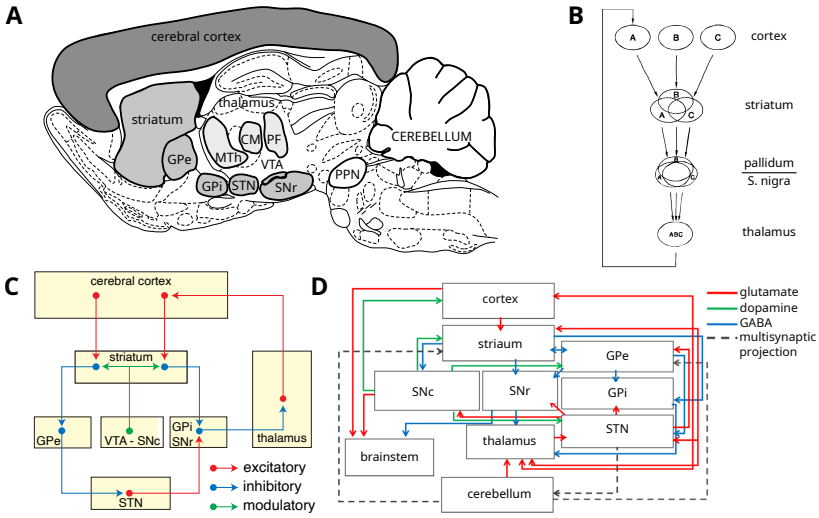


Figure 1.3: Evolution of descriptions of the cortico-basal ganglia-thalamo-cortical loop. A. Relevant regions involved in the CBGTC loop in a sagittal slice of the rat brain. The drawing highlights the cortical mantle, basal ganglia nuclei, motor thalamus and brainstem regions related to motor control. Adapted from Paxinos and Watson (1998) different shades of gray used to highlight cortex, BG and thalamic regions, from darker to lighter, other relevant motor controlling regions are also labeled. **B.** Canonical model of the CBGTC loop as presented by Alexander in 1986, where parallel segregated cortical processing units project in segregated but converging manner into the striatum. These in turn would maintain this layout with more overlaps in BG output regions SNr/GPi. Adapted from Alexander et al. (1986). **C.** An updated model of the canonical circuit, separating roles for BG direct and indirect pathways to control behavior, and assigning a modulatory role to signals from DA nuclei VTA and SNc. Adapted from Lanciego et al. (2012). **D.** More recent view of the circuit structures and their interactions, sharing reciprocal interactions that modulate their internal processes, maintaining the parallel & convergent organization. Adapted from Simonyan (2019). Abbreviations used in all panels are: GPe and GPi, external and internal segments of the globus pallidus respectively; STN, subthalamic nuclei; SNr, substantia nigra pars reticulata; VTA, ventral tegmental area; MTh motor thalamus; PF, parafascicular thalamic nucleus; CM central medial thalamic nucleus; PPN, pedunculopontine nucleus.

of these signals can converge in higher order somatosensory or motor regions to be used to guide behavior (Hamadjida et al., 2016). For example, higher order motor regions have been related to motor preparation in non-human primates and rodents, where particular activity patterns are associated to different movements (Bruce and Goldberg, 1985; Armstrong et al., 2009; Inagaki et al., 2018; Erlich et al., 2011; Boyd-Meredith et al., 2022). Notably, this higher-order agglomerations, also have a parallel organization themselves, which can be seen in the effects of local damage to small regions in visual, auditory, or somatosensory areas that can lead to particular types of agnosias (Hart, 2015).

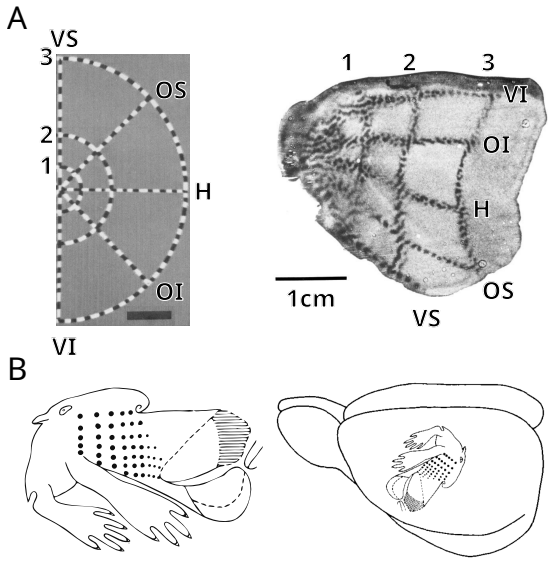


Figure 1.4: Cortical maps of two different sensory modalities. Different primary cortical territories receive organized patterns of projection from the sensory periphery. Cortical mappings do not need to be isometric to the input space, some regions of the periphery tend to be over- or under-represented. **A.** Stimulus (left panel) and resulting cortical activation in layer 4 of a primate brain by means of autoradiography (right panel), adapted from Tootell et al. (1988). **B.** Isolated representation of the rat body map, *ratunculus*, on the somatosensory cortex of the rat (left panel), and the approximate location of the body map in the rat cortex (right panel); adapted from Welker (1971).

Outside the representation of environmental elements, internal or external, by mean of sensory information integration. Higher order regions in the cortex are known to be relevant for decision-making and executive functions, and, as such, relevant for goal-directed behaviors (Friedman and Robbins, 2022). For example, the orbitofrontal cortex has been shown to have a role in prediction, action evaluation and confidence (Rudebeck and Murray, 2014; Masset et al., 2020), and fronto-parietal circuits show evidence of accumulation computations happening in the context of decision-making (Scott et al., 2017). These types of activity and the dense cortico-cortical projections allow the maintenance of rich representations of the state of the environment, and

the actions available in them (Cisek, 2007). Moreover, frontal regions have also been said to represent information about the circumstances that the agent or environment are in (Lee et al., 2012). In this way, higher order associative areas could allow the integration of abstract environmental information, such as goals, possible actions, and expectations of outcomes to regions receiving these signals. And, within this higher order associative areas, there is also evidence of parallelization and functional segregation, v.g. in the processing of different task phases (Kapoor et al., 2018), and this property might facilitate flexible behavioral control (Macpherson et al., 2021).

Importantly, this parallel, and hierarchical arrangement of the cortical territories, as shown by the organized mappings and architectural constrains of their divisions, is largely maintained in their corticostriatal projection (Nambu, 2011; Hunnicutt et al., 2016; Hooks et al., 2018), and in the patterns of projections from BG output regions to thalamus and from thalamus to cortex and re-entering the BG (Sakai et al., 1998; Yasuda and Hikosaka, 2018). We dive further into BG and motor thalamus, as they will be the main targets of the present work.

1.3.3 Anatomy of the basal ganglia

The basal ganglia are a group of interconnected subcortical nuclei, including the striatum, in primates the caudate and putamen; the internal and external segments of the globus pallidus (GPi and GPe, respectively); the pars reticulata of the substantia nigra (SNr); the subthalamic nucleus (STN); and two dopaminergic midbrain nuclei are also associated to it, namely the pars compacta of the substantia nigra (SNC), and the ventral tegmental area (VTA). It is a highly conserved structure across taxa, at large present in vertebrates (Stephenson-Jones et al., 2011), and with a likely origin dating far into the geological history of the planet (Grillner and Robertson, 2016). The input nuclei modulates the tonic inhibition exerted by the output nuclei over many target areas (McElvain et al., 2021), via direct suppression of their activity, or indirect excitation of them, leading to two complementary pathways. The BG has been largely studied in the context of voluntary motor control and inhibition of competing behaviors (Mink and Thach, 1993; Mink, 1996; Park et al., 2020), whilst currently it has been associated to a variety of non-motor behaviors, including decision-making, procedural learning and working memory (Wise, 1996; Seger and Spiering, 2011; Stephenson-Jones et al., 2016; Hikosaka et al., 2019). Dysfunction of these regions are associated, depending on the underlying etiology, to different motor, and or behavioral afflictions, ranging from Huntington's disease, Parkinson's disease, obsessive compulsive disorder, gambling, and drug addiction (Rapoport, 1990; Everitt and Robbins, 2005; Foerde and

Shohamy, 2011; Crittenden and Graybiel, 2011; Kalkhoven et al., 2014; Administration , US; Mestre-Bach and Potenza, 2023). BG nuclei are divided by their functional anatomy, where the input, striatum, projects to output areas, SNr and GPi, directly or mediated by intrinsic nuclei, GPe, STN. Dopaminergic signals from SNc, and VTA modify striatal and corticostriatal connectivity (Lanciego et al., 2012). Along the processing pipeline, BG nuclei tend to reduce their number of neurons, with striatum having the larger count of cells, $\sim$ 2.8 M; GPe & STN, $\sim$ 70 K; and finally SNr & GPi, $\sim$ 10K. From the output of BG, SNr almost doubles the number of cells in GPi (Oorschot, 1996).

1.3.3.1 Striatum, the organized input

Many cortical regions project to topographically segregated regions in the striatum (Hunnicutt et al., 2016; Hintiryan et al., 2016). These projections maintain the original topographical organization, where neighboring neurons in cortex, project to nearby cells in striatum, thus maintaining the original mappings (Romanelli et al., 2005; Nambu, 2011; Hooks et al., 2018), and these projections seem to maintain their lateralization. The striatum also receives projections from thalamic areas, in particular from the motor thalamus, including the ventral anterior and lateral nuclei (VA/VL), mediodorsal (MD), which land in striatal target receiving from motor and associative regions (McFarland and Haber, 2000). The anatomical organized cortico- and thalamo-striatal projections lead to functional divisions of the striatum. Generally, there are three main division recognized, the dorsal, ventral, and tail regions. In rodents, the dorsal striatum is subdivided into dorsomedial (DMS) and dorsolateral (DLS) regions, the primate caudate and putamen respectively. Where, the DMS receives orbitofrontal and limbic cortical afferences, and DLS receiving mainly prefrontal, somatosensory and motor projections (Hunnicutt et al., 2016; Lanciego et al., 2012). This anatomical organization of the projections drives to more value and action outcome signals in the DMS (Lau and Glimcher, 2007, 2008), and more motor related information in DLS (Crego et al., 2020; Cruz et al., 2022). The ventral striatum, includes the nucleus accumbens, core and shell, and mainly receives projections from limbic cortices and the amygdala. The dorsal striatum is associated with movement and habitual learning (Yin et al., 2009; Thorn et al., 2010), whereas the ventral more with value encoding. Finally, the tail of the striatum receives from sensory cortices and has been associated to avoidance and safety (Valjent and Gangarossa, 2021).

In terms of the cell population, the majority of the cells within the striatum are GABAergic medium spiny neurons (MSN) representing up to 90% of the population

(Gerfen et al., 2013). These are though to be the only cells projecting from the striatum, and have a very large negative resting potential, requiring the coincidence of many inputs to trigger an action potential. The MSN population is further subdivided into two classes, dependent on the dopamine receptor that they express. The MSN expressing the D1 receptor project directly to SNr, originating the direct pathway (dMSN), whereas the ones expressing the D2 receptor project to the external part of the pallidus and give origin to the indirect pathway (iMSN). Neurons expressing the D1 receptor, dMSN, increase their firing rate when DA is present, whereas D2-type receptor expressing neurons, iMSN, decrease their firing rates. On the one hand, dMSN suppress the inhibitory SNr/GPi, increasing the activity of tonically constrained BG target neurons. On the other, iMSNs can increase the activity of STN projections to SNr/GPi, by lowering the inhibition of the GPe, hence increasing the suppression on the receiving populations (Gerfen and Bolam, 2010; Verharen et al., 2019). It is considered that the inhibition and excitation of the output regions of BG, leads to an increase or decrease in overall activity (Tecuapetla et al., 2016); but more likely, their organized interplay is key in the process of action selection (Cruz et al., 2022).

1.3.3.2 The intrinsic nuclei

1.3.3.2.1 GPe, STN, inhibition and excitation inverted

The GPe and STN are relevant regions in the BG as they coordinate the indirect pathway, as iMSN do not project directly to BG output nuclei. Although, STN is also considered in an independent pathway—the hyperdirect, receiving direct cortical input to induce a fast-stop in general activity (Nambu et al., 2002), we are focused on the broader roles of these regions. STN, in contrast to striatum and GPe, presents a large majority of glutamatergic neurons projecting to the GPe and in tandem to GPi/SNr, thus it can directly and indirectly increase the inhibition of downstream populations (Parent and Hazrati, 1995b). Whereas, GPe GABAergic projections, are receiving striatal inhibition, or STN excitation, thus they can either increase the level of tonic inhibition exerted by the BG outputs, or decrease it (Parent and Hazrati, 1995b). Both regions receive topographically organized projections, that maintain the body maps organization present in motor and supplementary motor regions, whilst still presenting some level of convergence (Iwamuro et al., 2017). This organization could facilitate the role of BG in action selection and production of smooth behavioral plan. Supported by the effects of abnormal levels of dopamine in the BG, which leads to the parkinsonian symptoms by affecting the dynamics of these nuclei (Nambu and Tachibana, 2014). These control structures of the BG, through their mono- or multi-

synaptic projections into different regions, are able to influence BG output via their inhibitory or excitatory roles.

1.3.3.2.2 SNc, VTA, the dopamine in between

SNc and VTA, the sole sources of dopamine (DA) in the BG, project densely towards virtually all divisions of the striatum, and as most areas in the BG, they also have a topographical organization of the inputs (Joel and Weiner, 1994). DA is a key neuromodulator in BG, depending on the receptor type expressed by postsynaptic cells, the release of this amine can lead to an increase or decrease in the probability of an action potential. By mean of spike-timing-dependent plasticity, the release of DA can also modulate the internal connectivity in receiving regions, this is particularly relevant in striatum and cortex (Kreitzer and Malenka, 2008; Fields et al., 2007; Dehaene and Changeux, 2000). DA deficits have been long known to affect motor control, where, hypo- or hyper-active DA activity are associated to Huntington's or Parkinson's diseases symptomatology respectively (Crittenden and Graybiel, 2011; Florio et al., 2018). Nonetheless, aside from motor control, dysregulation of the DA signals, particularly in BG related nuclei, are also associated to maladaptive behavioral patterns and psychiatric disorders (Pallanti et al., 2010; Mestre-Bach and Potenza, 2023; Kalkhoven et al., 2014; Everitt and Robbins, 2005). Another role of DA activity have also been characterized in relation to rewards and cues, where stimuli that are predictive of rewards or unexpected rewards, increase the levels of DA cells activity once they are presented (Schultz et al., 1993). In case of an omission of the expected reward, a dip in the activity of the DA population is generally observed. This signal, named reward prediction error, relates a stimulus to an expectation of an outcome, and is a fundamental element in value based learning and reinforcement learning (Glimcher, 2011; Tobler et al., 2005). This role of DA activity can modify the corticostriatal connectivity, making the BG more likely to take actions that are expected to be valuable Shen et al. (2008); Balleine and O'Doherty (2010). Taken together, this evidence supports a role of DA terminals in the striatum in learning. For an agent interacting with their environment, actions associated to rewards are more valuable, hence should be more likely to be taken if available. In physiological terms, one can think that this kind of signals can help to guide action selection, as a more appetizing outcome could drive more activity in an ensemble of neurons, thus making some behaviors more likely to be taken.

1.3.3.3 GPi, SNr, continuous output regions

Despite their anatomical and developmental differences, the GPi and SNr are generally considered a continuous structure (Lanciego et al., 2012). Both consisting mainly of GABAergic neurons with high baseline firing rates that project to many regions in brainstem, cerebellum and thalamic territories (Gerfen and Bolam, 2010). Even though, in terms of afferences, GPi seems to receive organized motor projections, whereas SNr receive mainly orofacial somatosensory and motor projections (Nambu, 2011), and includes a region related to oculomotor control (Hikosaka and Wurtz, 1983, 1985). As such, the main modality of information processing for GPi would be segregated, whereas in SNr it seems to be based on convergence (Romanelli et al., 2005). In a similar line, DLS projects densely to GPi, whereas DMS projects to SNr, giving rise to the idea that both regions also enforce different levels of the BG computations. Namely, GPi communicating the motor, and SNr the cognitive components (Romanelli et al., 2005). Nonetheless, both regions have shown to vary their activity depending on the expected value of an action, although is more common to find SNr in this literature (Bryden et al., 2011; Sato and Hikosaka, 2002; Hikosaka and Wurtz, 1985; Hong and Hikosaka, 2008). And, in the same way, both are related to motor control, with GPi being more prevalent in this case (Romanelli et al., 2005; Basso et al., 2005). An interesting observation, is that VS projections targeting SNr alters M1 activity, thus presenting a level of convergence where limbic information can alter motor output (Aoki et al., 2018), as an example of an open loop portion of the CBGTC circuit. In terms of projections from BG outputs, both GPi and SNr project to motor controlling regions in the brainstem, including the PPN, superior colliculus, and others (Parent and Hazrati, 1995a; Gerfen and Bolam, 2010; Lanciego et al., 2012). And both also project to particular nuclei in the thalamus, the ventral anterior and ventral lateral (generally considered in conjunction VA/VL), the ventromedial (VM), and mediodorsal nucleus (MD). Where again, GPi seems to project to the nuclei that mainly project and receive to motor and sensory regions, and SNr to associative and limbic areas (Sakagami and Lattal, 2016; Aoki et al., 2018; Yasuda and Hikosaka, 2018; Kuramoto et al., 2011; Bosch-Bouju et al., 2013). The classical interpretation of their tonic inhibition over motor controlling region, indicates that once they are silenced an action is allowed to be taken (Hikosaka et al., 2000). Given the inhibitory role of the output regions, it might be surprising to find literature where increases in activity of them are related to vigorous activation (Sato and Hikosaka, 2002; Basso and Wurtz, 2002; Rizzi and Tan, 2019). But, we have to remember that it is important to select the correct action to achieve a goal, but maybe more important

is to not select competing actions. There are few correct ways of doing something, but many more of doing them wrongly. Given BG output regions position on the CBGTC circuit, they are the shortest path that BG computation have to reach cortex. With projections over thalamic nuclei which in tandem project to cortical and striatal regions in a topographically organized manner (Yasuda and Hikosaka, 2018; Bosch-Bouju et al., 2013).

1.3.4 Anatomy of the Thalamus

The last region within the CBGTC loop is the thalamus, in particular the nuclei associated in motor control, collectively denominated the motor thalamus (MTh). MTh is an umbrella term that encompasses the nuclei receiving and sending dense projections from motor, supplementary motor and associative regions of cortex, whilst also interacting similarly with the output of BG and cerebellum (Bosch-Bouju et al., 2013). In general, cortical regions project broadly to the MTh, with some level of organization, where motor and premotor regions project and receive from the ventroanterior and ventrolateral (VA/VL) nuclei, and associative regions targeting the ventromedial (VM) and VA nucleus (Sakai et al., 1998; McFarland and Haber, 2002). On the other direction, thalamocortical projections also present topographically organized mappings, with some amount of specialization in the layering of the targets in cortical regions, where motor and associative thalamic outputs reach different cortical layers (McFarland and Haber, 2002). BG and cerebellar targets in MTh are more segregated, with GPi and cerebellar projections reaching VA/VL, and SNr targeting VM and VA (Houk and Wise, 1995; Kuramoto et al., 2011; Çavdar et al., 2014; Hintzen et al., 2018). BG territories within MTh are mainly connected to associative and premotor cortices, which makes them an interesting target to study how movement preparation and expectations can be transformed into motor commands (Bosch-Bouju et al., 2013; Haber and Calzavara, 2009; Xiao et al., 2009). The vast majority of MTh neurons are glutamatergic, and project to layers I and II in cortex, and less to layer V, with few GABA interneurons, which still could be of relevancy in Parkinson's (Bentivoglio et al., 1991; Okoro et al., 2022; Albaugh et al., 2021).

Historically thalamic sensory regions have been classified as "drivers" or "modulators", by either increasing the drive of some population to relay information to cortex, or modulating the gain of some transmitted signal, respectively (Bickford, 2016). Given this background, this was also the case to MTh nuclei, but current anatomical and physiological evidence better supports the notion that MTh could process information differently than sensory nuclei (Garcia-Munoz and Arbuthnott, 2015; Worden

et al., 2021). Indicating that these areas could work as a "integrators" of cortical, BG, and cerebellar activity; relating cognitive and proprioceptive information to facilitate adaptive decision-making and learning (Jeljeli et al., 2003; Bosch-Bouju et al., 2013).

1.3.5 General organizing principles

According to the anatomical revision in the previous sections, we can observe that there are general organizing principles along the CBGTC loop. As a starter, along the loop, a parallel and hierarchical structure is maintained, where cortical territories and hierarchies are conserved in the receiving structures (Kim and Hikosaka, 2015; Hooks et al., 2018; Foster et al., 2021; Maurin et al., 1999). Even though, for the most part, information transverses the loop in separated channels, at each level, some amount of overlap, or convergence, is endured (Alexander et al., 1986; Miyachi, 2009; Nambu, 2011). Given this, at the level of BG output, SNr to MTh projections are in a great position to use limbic information to update motor controlling signals in cortex and BG (Aoki et al., 2018). A relevant feature that arises from these consistent mappings and general parallel organization, is that information about *what's going on* constantly re-enters into similar locations, this allows for recursion to be embedded within the system. This recursion offers, from one moment to the next, a manner to update behavior given previous experiences and expectations of the future. Moreover, this also makes more likely to take adaptive decision the next time, by the plasticity that DA signals endow into the system. This has led to the notion that the CBGTC circuitry could be sustaining multiple parallel representations, each accessing partial observations of the environment at hand, and using the DA signal to learn adaptive mappings (Lau et al., 2017). Another property from the circuit, is the characteristic pattern of oppositional signals happening throughout, chiefly present in the BG internal organization, the excitatory and inhibitory complementary signals seem to be a hallmark of the loop[2]. A classical view of these signals, assumes that inhibitory signals would reduce mobility, and excitatory would increase it; currently these two signals are considered more like an interacting system. In this latter scenario, execution of an action requires the inhibition of other competing plans, and the failure in this inhibition would lead to errors in the responses, by either initiating too early or executing inappropriate actions (Cruz et al., 2022). Lastly, the circuit at different levels seems to be engaging in some type of dimensionality reduction and latter expansion, akin to an encoder-decoder strategy. Where the large cortical projection of cortical state information, is projected to a smaller number of cells at the BG entrance; from there, the consistent reduction in number of cells in BG nuclei, could reduce this information

and route it through projections into MTh back into the cortex (Bar-Gad et al., 2003). In this interpretation, the BG output to MTh would be the informational bottleneck.

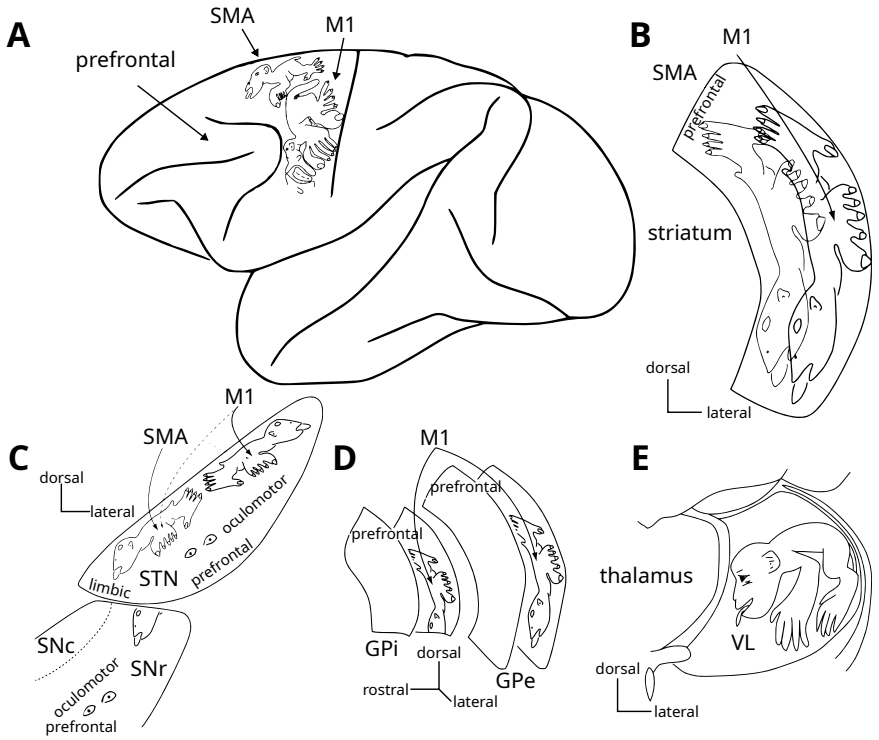


Figure 1.5: Overview of the CBGTC loop in the primate brain. A. Motor and supplementary motor cortical regions organize in body maps. Adapted from (Graziano, 2009), and Aflalo and Graziano (2006) citing Woolsey et al. (1952). **B.** Striatal, putamen, projections of cortical regions where the original mappings are maintained. **C.** Intrinsic nuclei STN presenting a full body representation, whilst SNr includes orofacial and oculomotor regions, aside from prefrontal projections. **D.** Intrinsic nucleus GPe presents a continuous map of the cortical maps, and GPi presents a mapping, missing the orofacial regions. **E.** VL nucleus of the thalamus is known to present an organized topography that sustains the original cortical structure. Images in panels B, C, D, E are adapted from Nambu (2011). Abbreviations included: GPi & GPe: internal and external capsules of the globus pallidus; M1: primary motor cortex; SMA: supplementary motor cortex; SNc & SNr: pars compacta and reticulata of the substantia nigra; STN: subthalamic nucleus; VL: ventrolateral nuclei of the thalamus.

Figure 1.5 depicts the essentials of the anatomical and functional properties of the CBGTC loop in a primate brain, as mentioned in the previous sections and the present. There is a general body-mapping maintained and projected along relevant portions of the CBGTC loop. Motor and supplementary-motor cortical regions (panel **A**), project to the striatum, particularly the DMS/putamen (panel **B**), and via the hyperdirect pathway into the STN (panel **C**). GPe and STN maintain the somatotopy

[2]Here we talk about excitatory and inhibitory signals in reference to the effects of the direct and indirect pathways on movement, not about the type of effect these routes have on downstream populations, which have the opposite effects.

(panels **C** & **D**). Whereas, the continuity of these mappings seems to break in the output areas of BG, with GPi maintaining the body up to the early portions of the face, and SNr receiving the orofacial section (panels **C** & **D**). Finally, in the motor thalamus, the body map is mainly observed on the VL nuclei (panel **E**).

1.3.6 What and how can BG communicate to the cortex?

If we ponder about the question *what can the BG communicate to the cortex?*, we can take into account the anatomical and functional organization of the projections revised. What are the consequences of the circuit level parallel and convergent architecture, recurrence, oppositional effects, endowed plasticity, and dimensionality reduction and expansion? From our revision, we can say that these allow the BG to receive cortical motor programs and movement parameters as proposals. Internally, BG can incorporate the limbic motivational components, and given the DA signals from previous experiences and habitual knowledge embedded into the synapses, it can select one plan of action. Via BG output regions, it can communicate this plan to MTh neuronal populations in motor, limbic and associative regions. In MTh, this selected action can be integrated with information about the current action in play, via the cerebellar proprioceptive information and error signals about movement execution. The integrated signal in MTh, can be sent back to cortex to facilitate the modification of the current plan, to prepare for a possible change in execution, or to maintain the current status. The general idea that cortical activity can be mapped into a dynamical action space, can be of help to image this process (Shenoy et al., 2013; Vyas et al., 2020). Within this framework, brain activity moves inside a manifold, where an action unravels as a path in neural population activity space. Given some initial state, the repeated execution of some actions with positive outcomes would facilitate the emergence of the same pattern. Negative outcomes, on the other hand, would imply the reduction of the probability of taking this action, and the need to explore new trajectories, until a good enough option were to appear. BG signals into MTh, could enforce the maintenance of the original trajectory, by focalizing the activity to facilitate the maintenance of a particular configuration, and selective inhibition of non-desirable actions. But, if a better option were to be available, or something were to change in the current environment, BG could reduce the inhibition in MTh, allowing it to increase the available space for exploration in cortical activity. This would give the BG a larger space of options from where to sample for better trajectories to select from. In summary, the basal ganglia can communicate information to the cortex to help select a motor plan that has been learned to be adaptive. This

communication prepares the necessary mechanisms for releasing the behavior when appropriate. Additionally, the basal ganglia can inform the cortex about the need to update the learned motor plan if task demands require it.

After considering the *what*, we need to consider the *how* question. More clearly: *how could BG communicate to cortex?* In all the presented descriptions of the internal connectivity of the CBGTC, the shortest path that BG evaluation process could take to enforce this kind of activity into cortex is through MTh projections. Modifications of a movement plan, or update an action in execution via brainstem or other motor controlling regions, would take longer to reach cortex. For agents that have to respond to ever-changing environments, this delay can translate into losing valuable time constrained resources or being captured by a predator. Another aspect of how BG could communicate to cortex, is the signals that it can send. Given expected outcomes of a selected plan, or an act in execution, BG has the ability to either leave it running, or—trough MTh—enforce the exploration of new proposals. As mentioned in the previous paragraph, by means of the inhibitory control over MTh population, BG output could maintain the status-quo, or by liberating the inhibition on selected populations, it could facilitate cortical exploration of new opportunities. In this sense, one of the most relevant elements of *what can be communicated* is that if no update is needed, there should be no relevant information passed. Only when something is amiss, unexpected, or a new opportunity arises, there has to be a change, as the saying goes: "if it ain't broke, don't fix it".

The general ideas outlined above are supported by experimental evidence that shows how particular channels of the CBGTC loop mediate movement control. Via BG output into thalamic territories that later would affect cortex (Inagaki et al., 2022), or including cerebellar nuclei (Wang et al., 2021; Schäfer et al., 2021).

1.4 Adding context to action-outcomes

The kind of tools developed to study behavior, as the ones presented in section 1.2.2, have been used to explore how animals can map arbitrary stimuli or actions to positive outcomes. For example, as commented in the aforementioned section, a tone, or a light could be used to inform an animal that a reward would be available, as in the classical conditioning experiments. But, it is also possible to chain the presentation of a stimulus with an operant behavior, v.g. pressing a lever after a tone to get a reward. This kind of assays give a tool to explore threshold levels of perception (Staddon and Cerutti, 2003). By changing the frequency, power, or signal-

to-noise ratio of the stimulus, and measuring the time the animal takes to make a response, reaction time, is a way of accessing the capabilities of the animal to detect those dimensions of the stimulus (Krantz, 2012). This paradigm can be made more complex, and train animals to select between different actions given some stimulus. For example, a reward can be given for a low frequency tone if is paired with left lever presses, and high frequency tones with right lever presses. After training, if the presented stimuli are drawn from values in between the trained low and high frequencies; one can find the value of the difference at which animals lose the ability to discriminate between frequencies, or other perceptual properties. In general, the use of operant behavioral boxes as helped to study sensory perception using the framework of psychophysics (Swets, 1961; Krantz, 2012; Akre and Johnsen, 2014).

An important property of the aforementioned paradigms, is that all correct actions in these assays lead to reward. If one were to relate brain activity to these responses, it would always include some form of reward expectation, confidence or other information related to the outcome (Masset et al., 2020; Klein et al., 2012; Oswal et al., 2007). To better isolate these motivation components from action preparation and execution; Reiko Kawagoe, Yoriko Takikawa, and Okihide Hikosaka developed an oculomotor task for monkeys. In the assay, animals were asked to make eye movements to lateral targets, but in blocks of trials only one of those targets would be rewarded, the 1-direction rewarded task (1DR) (Kawagoe et al., 1998). At the end of the block, the rewarded target would be changed without informing the animal. In this way, animals had to make movements to the same targets, but with different expectations about the result, separating the reward expectation from the movement preparation. But, importantly, it also gave the animals the opportunity to learn that there were circumstances that gave the same actions different values. There was a context that they could infer that implied that one target was rewarded and not the others. This seminal paper, showed that activity of neurons in the primate caudate, DMS, is modulated by the expectation of rewards. It also showed that behavioral correlates of the movements were different depending on outcome expectation: rewarded movements were initiated more avidly than non rewarded. This behavioral paradigm has been developed into different versions, the original one presenting four possible targets, with the presentation of the target cue briefly during the fixation period, but before a go cue, Fig. 1.6 top; but, other version presented only two possible targets, and have variations of the relationship between the target and go cue presentation, Fig. 1.6 bottom.

The main behavioral results from the original and variations of the 1DR paradigm are consistent, the different tasks versions consistently drive animals to respond dif-

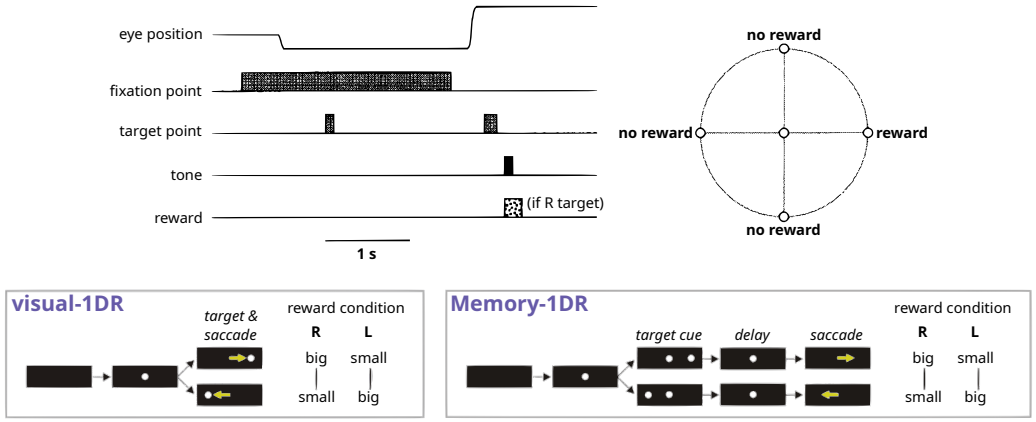


Figure 1.6: Visual depiction of different versions of the 1 direction rewarded (1DR)tasks. On the top, Original version of the 1DR task with four possible targets, and only one rewarded in each block. The target point was presented after a 1 −second delay from the moment the monkey fixate in the center fixation, after a variable delay, the fixation point would disappear and the monkey had to saccade towards the cued location. On the bottom, two variations of the 1DR task, with only two targets. In the visual-1DR, bottom-left, the target and go cue are presented simultaneously; whereas, on the memory-1Dr, on the bottom right, the target cue is presented briefly after the fixation, and the animal has to remember the location to saccade into that place after a go cue. In both cases, there are two independent and iterated reward-schedules, in one case rightward movements receive a big reward, and leftwards movements a small one; or the opposite. Panel on top adapted from Kawagoe et al. (1998), panels at the bottom are adapted from Hikosaka et al. (2006).

ferently towards the targets depending on their expected outcomes (Fig. 1.7). When animals are asked to make a non rewarded movement, they initiate the movement slower than for rewarded trials (Fig. 1.7 A), implying that during the task animals are aware of the different context that they are experiencing. The simplification of the task into just two targets (Fig. 1.7 B & C), allowed the researchers to gain deeper insights about the behavioral effects after a block transition. For starters, it became clear that animals were initiating the blocks unaware of any difference with the targets values, (Fig. 1.7 B & C), initiating movements quickly for previously rewarded, now non-rewarded, targets; and, initiating slowly towards the previously non rewarded, now rewarded trial. But, what happened in the behavior after the first trial was also quite indicative that animals in fact had a notion of the different context in the task. When they analyzed the data separating responses by trial type within a block (Fig. 1.7 B), there was a noticeable difference in the way responses changed towards the currently rewarded or non-rewarded target. Animals took longer to update their responses towards non-rewarded locations than towards rewarded, observable in the number of trials animals took to display RTs compatible with the new reward locations. Moreover, when looking closely at the first two trials after a transition and separating trials by target value (Fig. 1.7 C), researchers noticed that responses up-

dated for both the previously experienced and non-experienced targets. This result supports the notion that animals were aware of the two different context, and were using this information to infer changes in the context from a change in one response.

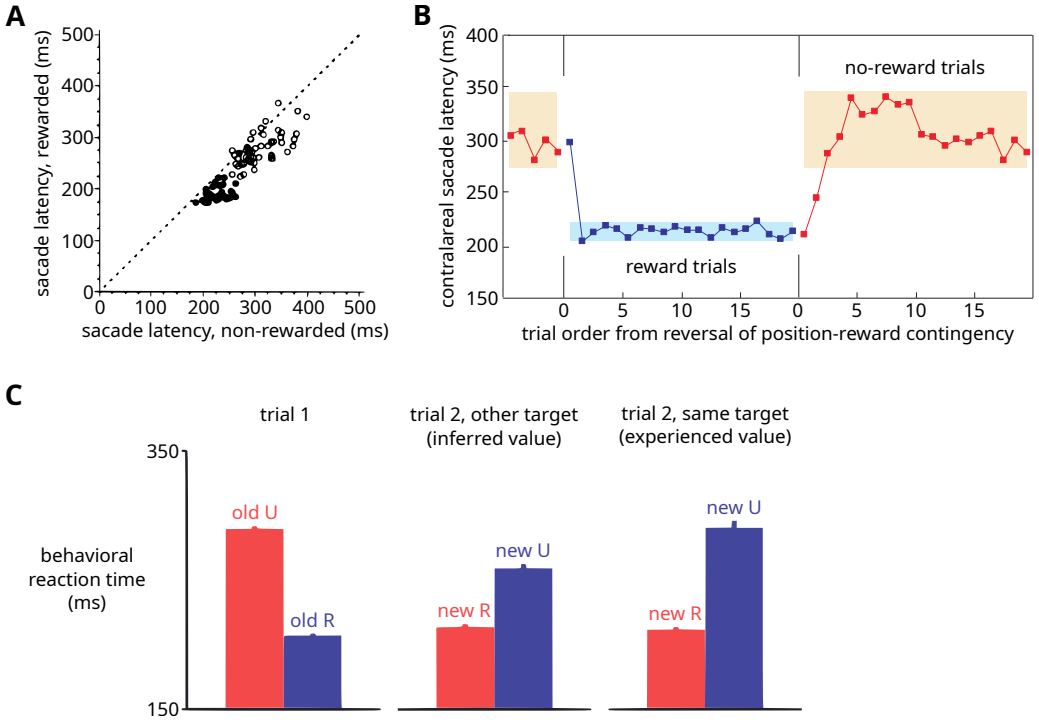


Figure 1.7: Different versions of the 1DR task lead to consistent results. **A.** In the original publication of the task, two monkeys were trained and tested in the task for a number of multi block sessions. The average latency to initiate saccades towards the non-rewarded side was longer than for rewarded targets. Each dot represents a session, and the circle filling represents the individual animals, adapted from Kawagoe et al. (1998). **B.** Behavioral results from two monkeys in the visually guided 1DR, showing average saccade latencies for rewarded and non-rewarded trials after a block transition separating by updated target value. For the first trial after the transition, saccades initiate as they had previously, quick for previously rewarded, slow for previously non-rewarded. But, after the surprising outcome, they quickly update their responses for rewarded trials, whereas for non-rewarded targets it takes them more trials to initiate slower. Adapted from Lauwereyns et al. (2002). **C.** Behavioral results from another visually guided 1DR, showing reaction times, saccade latencies, after block transition, conditioned on first and second trial target. As in B, for the first trial, responses are consistent with the previous block mapping, but in the second trial, animals update their behavior for both experienced and non-experienced targets. Adapted from Bromberg-Martin et al. (2010b).

In line with the behavioral results, the neural correlates of brain regions that were recorded during the different experimental sessions drove further our understanding of their roles in the preparation and execution of eye-movements (Hikosaka et al., 2006). In their model, cortical information about stimulus identity could inform caudate receiving neurons about laterality of the target. The striatal neurons, depending on the history of DA signals from SNc/VTA, would project to SNr and facilitate or bias

the movements towards the rewarded location. Later work from the same laboratory has also shown that different populations within the caudate, DLS, encode stable or flexible value mapping. One population receiving prefrontal projections, would be in charge of flexible mapping, and the other receiving mainly from the temporal cortices care about the stable values (Hikosaka et al., 2014). And finally, the same group has also shown how motor thalamic neurons in BG receiving territories encoded stable or variable object values (Yasuda and Hikosaka, 2018). The sum of these results, gives an overview of how the CBGTC loop could be modulating behavioral responses with respect to outcome expectation. The modified reward schedule of the task, permits the characterization of how a motor target value can affect the internal processing of a plan. Where, contextual information, interacts with a visual signal, that inherits a value from their expected outcome. This integration facilitates or biases the selection of the movement towards the reward paired location, whilst making it more difficult to initiate the non-preferred one.

In summary, the 1DR task allows observing how the underlying circumstances that embrace a situation, can map into differences in the response patterns to the same stimulus. Thus, allowing to map not only actions to outcomes, but how similar actions in different contexts map to different outcomes. The implementation of a rodent version of this task could facilitate the study of all of these relevant phenomena in a more tractable animal model. Even thought rats do saccade to hold their gaze (Chelazzi et al., 1989), a freely moving version of the 1DR task would allow them to express their intentions in more open ways. Finally, this would allow the use of genome editing tools (Shevtsova et al., 2005; Chenouard et al., 2021), to further characterize the cellular populations responsible for the behavioral phenomenology in different brain regions

1.5 Objectives, questions, and hypothesis of the present study

The world presents as a collection of events in particular circumstances. We learn that something about the underlying setting in which we experience these particular states, can inform us that some actions are better than others. In general language, we use the word “context” for these elements that give meaning to the mapping between states of the world and actions-outcomes. From repeated experience, we form and learn these contexts. Thus, knowledge learned through experience, allows having a mapping of actions and results for particular states of the world that are flexible

to the underlying setting of the current state. With these, when confronted by a novel environment whose underlying circumstances share familiar resemblance to some previously experienced ones, we can use this prior knowledge to prepare actions given our expectation about their outcomes. In a sense, this boils down to Wittgenstein's proposition 1.1 in the *Tractatus Logico-Philosophicus* "The world is the totality of the facts, not of things" (Wittgenstein and Ogden, 1999). We do not experience the world directly, but our interpretation of what is out there given what we have learned, and where and how we find ourselves.

After giving the general overview of the relevant behavioral and neural mechanisms generally associated to context guided behavioral control, we can present the main and specific objectives of the present work. These objectives drive a set of question that put to test our hypothesis about how the brain can commit or modify a response depending on the context that he finds himself in.

Our main objective is to understand the way in which the BG can modify cortical activity to guide adaptive actions selection. To this end, we first need to have a task that requires flexible action outcome contingencies to depend on contextual information. Such a task would engage the relevant circuits within the CBGTC loop, to allow us to characterize the relevant elements in the circuit in the context of the task. In particular, the BG output plays a crucial role in influencing the cortex through MTh projections. The substantia nigra pars reticulata (SNr) and the ventroanterior and ventrolateral (VA/VL) thalamic nuclei are primary candidates for facilitating the maintenance or updating of a particular action plan. SNr receives convergence of motor, limbic, and frontal cortical information, which is then projected to territories of the MTh with efferences to frontal and motor cortical areas. Taking the aforementioned into account, to advance our general goal, we devise the following specific aims: (1) develop a rodent version of the 1DR task; (2) characterize the behavioral effects of manipulating reward schedules; (3) describe the neural correlates of these behavioral effects in the activity of SNr and VA/VL thalamus; and, (4) relate the activity of these regions to the process of commitment or update of a motor plan. The questions that arise with respect to objectives 1 & 2 are: (I) Do rat behavior recapitulates the primate results?; (II) Can we modify the 1DR task to gain further insights about movement preparation and context representation?; (III) Are rats capable of inference-like behavior in the task?; (IV) Does the freely moving aspect of the task allow animals to express something about their knowledge in their behavior? From the information presented in our revision of the current state of the research field, our hypothesis in relation to these questions are the following: (A) Given the pervasive role of vigor in motor control, rats should also recapitulate in general the primate results;

(B) Given effects of predictable delays on reaction times, we expect that by adding variable delays between the fixation, movement and go cue presentations, we will gain leverage about the effects of reward localization and behavioral biases; (C) If animals were to develop context-dependent preferred actions, they should also be capable of developing inference-like behaviors. As a surprising result implies that the context has changed, and the other target has also changed in value; (D) Animals could embody the knowledge about context in their behavior in different ways, the simplest would be to initiate trials in different orientations depending on the context they know to be in. Thus, we expect animals to initiate trials orienting themselves differently to facilitate initiation of movements towards the rewarded target. With respect to objectives 3 & 4, research questions that arise are: (I) What are the kind of signatures represented in SNr and MTh? (II) How do these signals evolve at the single neuron or population levels during the task? (III) Are the population level signatures stable over time? (IV) Do these signatures include relevant properties that could facilitate the commitment or update of a default motor plan? The anatomical and functional properties of the recorded regions, allows us to hypothesize that: (A) The regions could present signatures about movement direction, expected target value, and context of the current trial; (B) Given the task block-structure, single neurons and the whole populations could be informed about context before trial initiation, about movement direction at movement cue presentation, and about value once these two signals were to interact; (C) As both regions integrate motor and limbic components, during the delay before go cue presentation, population level signatures in both regions should present stable structure for direction and value; (D) SNr signals could inhibit MTh populations, to allow MTh to not enforce these particular actions. Whilst, MTh could modify cortical signals by facilitating the update of a behavioral plan.

In this work, we treat objectives 1 & 2, with their respective questions, and evaluate the respective hypotheses in chapter 2. We do the same for objectives 3 & 4 in chapter 3. Finally, chapter 4, presents the general discussion of the results, and how the bulk of the presented work relates to the main objective, discussing also the limitations, and presenting future directions to be explored.

Chapter 2

Behavioral signatures of context in a delayed movement task

I see you shiver with antici.....
pation

Dr. Frank N. Furter

2.1 Introduction

When speaking about context, we generally mean a particular setting that allows us to give meaning to a state of the world. Contexts allow understanding that similarities between states mean something for action selection. What we have experienced in the past simplifies the assignment of what can be done in the future. Over a lifetime, we learn that different context map similar actions to different outcomes, and that some actions are better to be taken than other depending on our previous experiences. Similarly, we learn that contexts can vary, such that when we experience an unexpected outcome in a known context, it means that something has changed; thus, our mapping must do so too. These grounded previous experiences about which actions are related to better outcomes, allows subjects to bias the selection of actions once they recognize the context that they find themselves in. For example, we can think of a goalkeeper in a soccer match that is trying to save a penalty kick. If the goalkeeper knows about the kicker's tendencies, he could prepare to lunge toward their preferred shooting direction, thus increasing the chance of success. If the shooting player were to act differently during his preparation, or there were a change of player taking the kick, our goalkeeper needs to update his initial plan. The time this takes will affect his ability to catch the ball. This example illustrates how being prepared allows us to make the most out of situations that we know how to navigate,

but that these mechanisms need to be flexible, to allow for adjustments or updates if the context changes. This is crucial in a world that is in constant development.

For animals, who do not have to catch penalty kicks in their ecological niche, being aware of the context that they are in allows them to behave and forage in more advantageous ways. In a food rich patch, recognizing the smell of a kin or a predator, will prompt them to react in different ways. The underlying machinery necessary to map external evidence to appropriate actions, has to be flexible to environmental changes, to allow animals to map context, states, and actions relations. A rich tradition of experimental psychology has shown that once an animal acquires an understanding of action-outcome contingencies, in a consistent and stable environment where cues are unambiguous, their responses become faster and more stereotypical (Thorndike, 1898; Welford, 1986). These responses do not need to have a causal relationship with the outcome, but only to appear relevant for the subjects. As in the superstitious pigeons of B.F. Skinner (Skinner, 1948), or in the stereotypies that animals develop while preparing a delayed movement in a timing task (Kawai et al., 2015).

Previous work has shown the relevance of outcome expectation in the production of movements. For example, the amount of effort that animals are willing to make to obtain the same amount of reward is related to the reinforcement value (Hodos, 1961; Sclafani and Ackroff, 2003). Vigor, the amount of effort to exert in the initiation and execution of an action, has been shown to depend on the expected outcome of an action (Choi et al., 2014; Shadmehr et al., 2019), pointing to the fact that not only the production, but the initiation of movements are affected by these anticipated outcome expectations. Reaction time (RT), or sometimes response time, in an instrumental task, is the period between an imperative stimulus and the initiation of a movement. This variable has been used to study how mental processes guide decision-making (Galton, 1890; Donders, 1969). Not only stimulus strength, sensory modality, difficulty of the response, or level of practice can influence RT distributions (Froeberg, 1907; Bertelson, 1967; Henry and Rogers, 1960; Welford, 1986). The delay period given for preparation, has been shown to also impact RTs, where in simple RT tasks longer time to prepare are associated with shorter initiation of movements (Nickerson et al., 1969; Henderson and Dittrich, 1998; Hackley, 2009; Vallesi et al., 2014). RTs tend to increase with normal aging, by affecting movement preparation more than initiation (Hardwick et al., 2022); And, the effects of preparatory cues are also affected in aging, Parkinson's disease and cognitive decline (Jurkowski et al., 2005; Mioni et al., 2018; Capizzi et al., 2022). Importantly, in our context of interest, namely: how the prospect of a reinforcement drives changes in performance? The expected outcome of the upcoming action has been shown to bias the response-initiation time (Calaminus

and Hauber, 2009; Bundt et al., 2019; Milstein and Dorris, 2011; Takikawa et al., 2002).

To study how knowledge about context could affect the execution of an upcoming movement by having different motivational value, the laboratory of Okihide Hikosaka developed a primate oculomotor saccade task. The key manipulation within this task was the value of the targets (Kawagoe et al., 1998). Initially, animals were required to fixate in a central located visual stimulus; after a fixed delay, one of four possible eccentric targets would be presented briefly, indicating the location to where the subject was to later make a saccade on that trial. Finally, after a second variable delay, the fixation stimulus was extinguished, cueing the animal to saccade towards the previously cued location. There were two reward schedules in this protocol, one where every correct saccade was rewarded, and another where only one of the target locations would receive a reward, "all direction rewarded" (ADR) or "one direction rewarded" (1DR) respectively. If animals were to abort the trial or not execute the requested saccade, the trial was repeated. In the 1DR protocol, after 60 correct trials, the rewarded target would randomly switch to a new one without informing the animal. This simple manipulation allowed to enrich the understanding of the roles that primate caudate neurons have in the production of saccadic movement and how their activity vary depending on the expected outcome of the targets (Kawagoe et al., 1998). The original and variations of the task gave the lab a tool to assess how different levels of the pipeline controlling the oculomotor response are affected by, encode, and communicate the expected value of the upcoming movement (Hikosaka et al., 2006, 2000, 2014). This circuit includes cortical, subcortical, cerebellar, and brain stem regions; and, in their interpretation of the results, the cortico-basal ganglia-thalamo-cortical (CBGTC) loop was the key circuit underlying the decision of the motivational value and salience of the expected outcome, and the preparation and updating of a movement towards specific targets. Where information about value and salience would be encoded and saved by dopaminergic modulation of the synaptic strengths of a neuronal population that would relate a particular state of the environment to the optimal actions (Bromberg-Martin et al., 2010a; Nakahara and Hikosaka, 2012). In the work with non-human primates using variations of the 1DR-ADR, the behavioral results consistently showed an effect of the expected outcomes in the execution of movements, depending on expected reinforcement value. In particular, animals persistently initiated saccades faster for rewarded in comparison to non-rewarded targets, and after a change in the action-outcome contingency—a transition—animals updated their responses (Kawagoe et al., 1998; Lauwereyns et al., 2002; Takikawa et al., 2002; Ding and Hikosaka, 2007; Matsumoto and Hikosaka, 2007). Importantly, after a contin-

gency change, monkeys update their responses globally, not only to the experienced target. Implicating that primates understood that the task had only two discrete contexts and used this information to map changes in environment to changes in their behavior (Bromberg-Martin et al., 2010b). This has also been observed in tasks with similarly discrete action-outcomes contingencies (Saez et al., 2015).

Here, we develop a rat version of the 1DR task. This assay provides a way to explore in a tractable model organism the manner in which context modify responses towards the same targets, as their value is changed in a structured manner. In our task, targets are nose ports, and context corresponds to the port where reward would be available if animals were sent to that location after fixating their snouts in a central port for enough time. Importantly, as in the primate version, animals are asked to move to both rewarded and non-rewarded targets. And after a number of valid responses, the rewarded location is flipped without informing the animal (Fig. 2.1 A, for an in-depth description of the task and training procedure see methods 2.4.3). To better isolate the effect of context in the preparation of actions, we remove the memory component, maintaining the presence of the target cue location after the first delay. In addition, we vary the delays between all cues. Both delays that animals will be required to wait while fixating at the center port are sampled from truncated exponential distributions, flattening the hazard rate of their arrivals. Thus, enforcing that animals need to pay attention to the sensory cues and not fall into a pattern of simply timing their responses.

Given previous work on similar behavioral tasks (v.g. Kawagoe et al. (1998)), we expect animals to respond differently depending on the expected outcomes. This could be observed in animals initiating movements quicker for rewarded locations with respect to non-rewarded, or being more likely to break fixation after movement cue presentation for targets that are not going to be rewarded. Initiating movements differently towards the rewarded and non rewarded location could appear in two possible ways in the RTs distributions, as either a shift or a scaling of RTs distribution. Explicitly, these effects imply that they could initiate movements later but with similar distributions, or have differences in the underlying distribution of RTs respectively. The aforementioned observations are in line with movement-vigor effects, where, rewarded movements drive more energy or attention than non-rewarded ones (Shadmehr et al., 2019). But, due to the maintenance of the movement cue during the go cue delay, and the variable duration of the latter in our paradigm, we hypothesize that if animals are aware of the context, they could develop a default motor plan to go towards the rewarded target. As this would make the action selection process more efficient. We reason that if animals were to prepare a default plan in advance of the

movement cue, we should expect that (1) for broken fixations before the movement cue presentation they will be more likely to go to the rewarded location; and (2) that the duration of the go cue delay will be negatively correlated with the animal RTs for the non-rewarded targets, but have negligible effect for the rewarded ones. These are implied because a default plan to go to the rewarded target supposes a facilitation of movements towards the rewarded location, therefore if a trial is aborted before being informed about target direction they should initiate their default plan. And secondly, taking into account a loaded plan to go to the rewarded target, movements cued to that location should initiate with similarly short RTs, but for movements cued towards the non-rewarded target, animals need to update their movement plan. This latter point specifically implies that: for short go cue delay durations, as animals have little time to update, leaving the center port should be slower, whereas for long delay durations they had time to prepare the new movement, hence can be quicker. An important consequence of animals using context and not just experience to guide their behavior, is that by keeping track of context, they could update their responses globally after a surprising outcome, consequently changing their behavior towards the non-experienced contingency. Finally, given that in our task are animals are freely moving, they might also facilitate the initiation of their preferred movements by initiating trials with different orientations depending on their preferred target. Thus, embodying the default motor plan and tracking of the context that they believe to be in.

2.2 Results

2.2.1 Target values changes reaction time profiles

In the delayed movement task (Fig. 2.1 A), the values of the lateral targets are flipped multiple times within the session in a block structure. The behavior of the animals showed that within blocks (Fig. 2.1 B), after a transition in rewarded direction, animals produced different response times depending on the expected outcome of the targets. They responded quickly for rewarded movements, and more slowly for non-rewarded movements. These differences in RT were sustained up to the next transition. Across animals, from the fifth trial after a transition, the RTs for rewarded and non-rewarded targets stabilized. Furthermore, the rate of changes for response time in trials after context transition differed between rewarded or non-rewarded trials. They quickly adapted and initiated with fast RTs for rewarded trials after the first unexpected outcome, whereas animals were slower to adapt their RTs to the

previously rewarded, now non-rewarded movement direction. This suggests, as has been shown in previous studies, that animals learn quicker from rewarded outcomes (Lauwereyns et al., 2002). We initially focus on stable trials, where the contingencies are clearly driving the animal's behavioral responses. In this regime, we calculate the hazard rate for broken fixation after the movement cue was presented, splitting trials by target cue value (Fig. 2.1 C), to estimate the probability of leaving the center port during the go cue delay. In this case, we notice that briefly before any go cue could be given (before 250 ms, shaded area in the panel), animals have a higher probability of breaking fixation for rewarded trials. But, from that moment onwards, animals are more likely to leave the center port for non-rewarded movements. This is indicative that animals prefer to abort trials that will not lead to reward.

From these observations, we can expect that at least two possible effects could drive the differences in reaction times and the expected outcome. On the one side, it could be the case that responses for non-rewarded targets are shifted with respect to the rewarded responses, or they could be scaled (Fig. 2.1 D). If we were to split the reaction times of each animal into quantiles depending on the expected outcome of that trial, we could draw a line connecting these values in a 2-dimensional plane. Here, a diagonal line would imply that both conditions have equal values, which we know is not the case. However, the connecting line could present at least two different linear transformations: it could either have a difference in the intercept, which would appear as a shift from the origin, or it could have a different slope, implying a scaling of the responses by some factor. The shift implies that reaction times are generated by a similar underlying process, but that some intermediate step is moving the starting point from the original one. Scaling, as a change in the slope of the distribution, implies a variation in the internal processes guiding behavior generation.

First want to evaluate if there is an effect by either target direction or value, for this we split the RTs data by subject and compare the difference between subject medians by either target direction or value (Fig. 2.2 A & C). The differences between median RTs per subject by target direction showed no significant effect under a t-test (mean= 0.178ms, $STD = 44.114$ms; $t(13) = 0.015$, $p = .988$), whereas the difference between median RTs per subject by target value has a highly significant difference under the same test (mean= -159.107143ms, $STD = 117.779974$ms; $t(13) = -8.10922024723809$, $p = 9.635 * 10^{-07}$). These results show that the difference between distributions comes from the target value, not the movement direction. Secondly, we wanted to evaluate if the difference between distribution was related to a shift or scaling of the underlying distributions. For this, we first split the distribution of reaction times for each target value into quantiles, and compare the differences

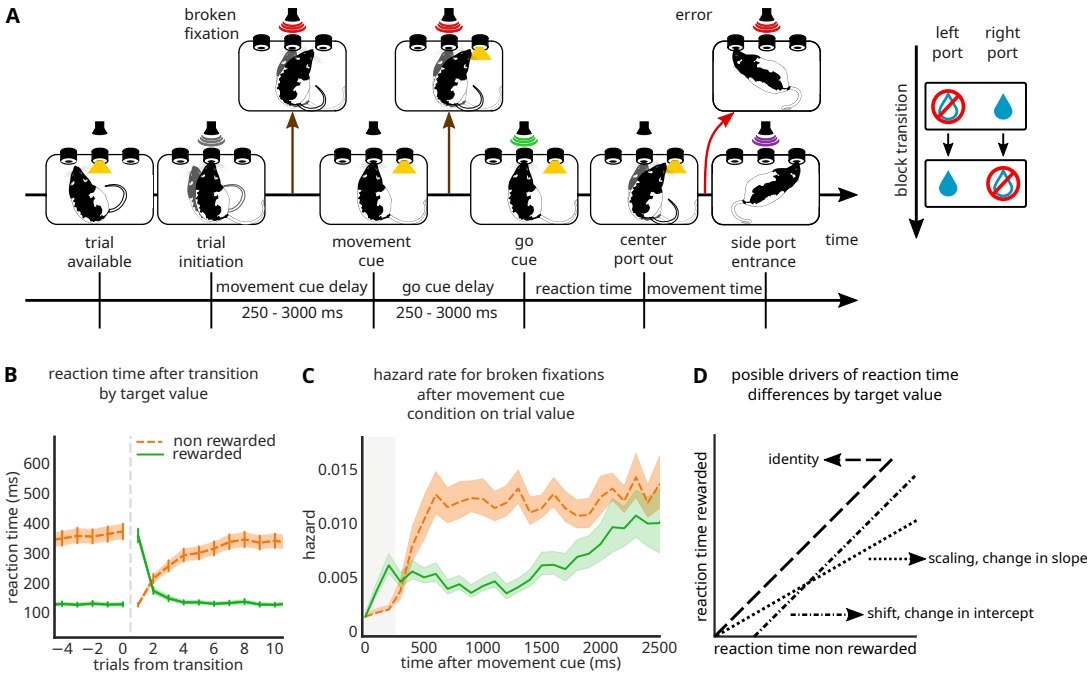


Figure 2.1: Task graphical depiction, stable trials selection and putative drivers of effects by value. A. Graphical depiction of the task: Animals are asked to wait while fixating in the center port until a *go cue* tone is given. They are asked to make pseudo randomized movements towards the lateral port that has his LED lit. Importantly, for variable numbers of trials, only one port gives reward (detailed description of the task and training procedures, in the methods 2.4.3.3). **B.** Effect of the block transition in animal reaction time by target value: The task reward contingencies and block structure scheme have different effects in animals behavior. Each animal contributes with their median reaction times for each condition across all their correct and valid trials for all valid sessions. Colors and line styles depend on target value, dashed and *orange* for non-rewarded and continuous and *green* for rewarded. The mean and SEM estimates are ordered by their position with respect to the last transition, the first block of each session is removed from all analyses. Any further analysis only considers trials after the fifth since a transition has occurred, unless otherwise noted. **C.** Animals are more likely to break fixation for non rewarded trials: The panel shows hazard rates for broken fixation after movement cue presentation by target port value, we calculate the hazard rates for each animal independently, and present the mean and SEM across them. **D.** Possible drivers of the reaction time effects in animals behavior: If reaction times were to be split by target values and in quantiles, and depicted using for each point their values by condition as their coordinates. The target value could affect the reaction times distribution in at least two possible manners, by either shifting or by scaling it. In the first case, this would appear as a change in the intercept of the slope of the reaction times regression line; in the second, there would be a change in the slope of the line.

between distribution depending on target value. When comparing the differences between means across quantiles, by session in the single animal or across animals, panels A & C in figure 2.2, respectively. Using a Wilcoxon signed ranks test, we observe a highly significant difference for all quantiles ($Z = 105$, $p = 6.103 * 10^{-5}$ for all quantiles independently). We then take these distributions, and use their mean over session medians for the example animal (Fig. 2.2 B) or the means over single-animal medians for the population (Fig. 2.2 D) as 2-dimensional coordinates. By doing so, we observe that the putative line that would connect the quantiles presents a slope change with respect to the identity line. Specifically, the quantiles line has a shallower slope than the identity when placing the rewarded trial RTs on the Y axis. This implies, as it is noticeable in figure 2.2 A & C, that non-rewarded trials have slower reaction times than the rewarded target trials. But it also implies that the distribution of reaction times for non-rewarded trials are scaled with respect to the reaction time for rewarded targets. These effects are consistent with reward affecting vigor through the application of a gain factor to the temporal dimension of the process underlying movement generation (Pardo-Vazquez et al., 2019).

These types of effects are generally associated with differences in the motivational value of the expected outcome, where animals are willing to expend more resources to generate a quicker movement towards targets that have a higher value. The behavioral component modified here would be the amount of effort to exert in movements depending on expected outcome, what is called vigor (Yoon et al., 2018; Shadmehr et al., 2019). But, aside from vigor driving these differences, other processes unrelated to the initiation and execution of the action could underlie these effects, such as a default motor plan.

2.2.2 A default motor plan driving the behavioral differences

Once animals are aware of the current context, they might have a default action prepared. Particularly, they might plan to go to the port where the reward could be given in each different context. To evaluate this possibility, we first look at how animals behave when they incur in a broken fixation, leaving the center port prematurely, before the movement cue is presented (Fig. 2.3 A & B). Before the movement cue is presented, animals have no information about where they will eventually be sent. A tendency to choose a particular site implies that they have a bias to go towards that side. We look at the probability of going to the port where reward could be given after broken fixation (Fig. 2.3 A), and compare the observed probability against a null hypothesis that they have no preference. The analysis shows that animals

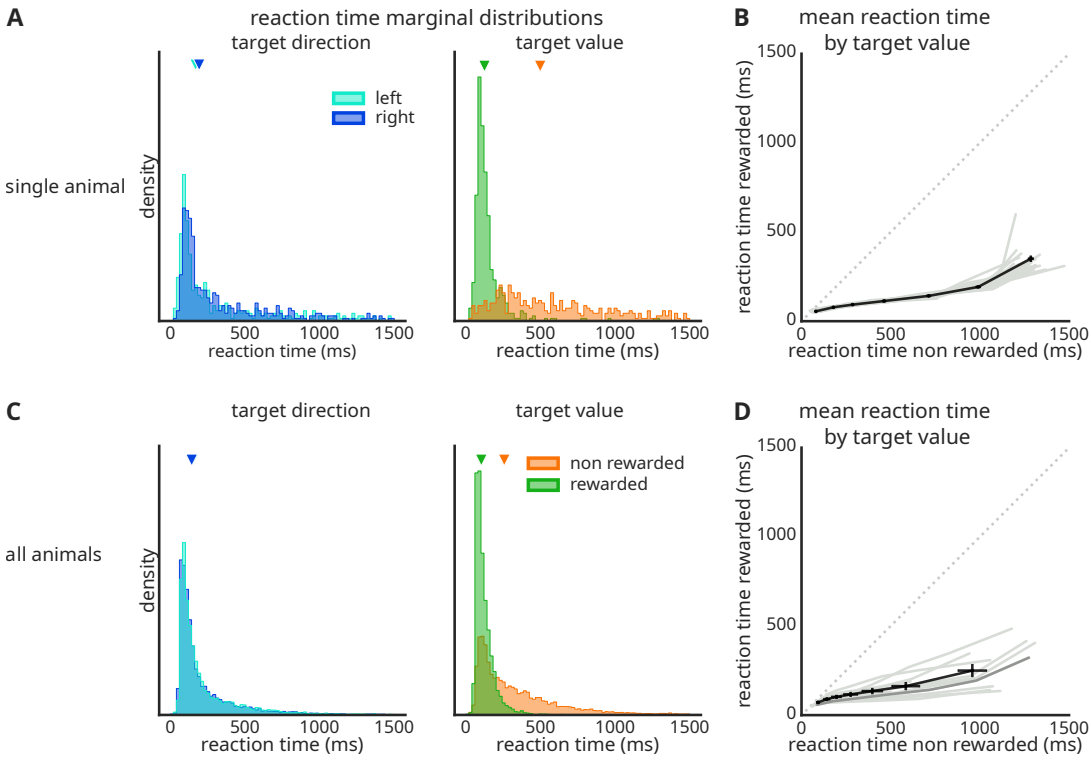


Figure 2.2: Reaction times distributions are unaffected by target direction, but scale by target value. A. Representative animal shows no effect by target direction, but clear effect by target value: In the marginal distribution for all valid and correct trials for all session of the example animal, each session contributes with the median reaction time for all block within the session. The trial results are marginalized with respect to their target direction, in the left, or value, in the right. The color encoding used for target direction maps left to *cyan*, and right to *blue*. The mapping for target value color non-rewarded in *orange*, and rewarded in *green*. **B.** Representative animal reaction times have scaled distributions: Median reaction times for all session (*gray* lines) of the same example animal, valid and correct trials within the session are split by target value and their reaction times ordered in quantiles. The *gray* lines connect the medians between quantiles per session, the values for non-rewarded trials are used as abscissa and for the ordinate we use the rewarded. The *black* line is the mean quantiles across session for the animal, and error bars represent SEM across session by quantile per condition. **C.** Same as A for all animals. In this case, for each animal, all valid correct trials for all sessions RTs are split by target condition and their median by block number is calculated. Each animal contributes with their 2 values per block numbers over sessions. **D.** Same as C for all animals. All valid correct trials per animal are split by target value, and split in quantiles. We plot the median value of the RT for each quantile across session per animal, each *gray* line, the lines connect the median RT for non-rewarded as abscissa and rewarded as ordinate. The *darker gray* line is the example animal presented in the previous plots, and the *black* line is the mean across animals.

are more likely to decide to go to the port where reward would have been given ($mean = 84.216\%$, $SD = 9.273\%$; $t(13) = 13.807$, $p = 1.91 * 10^{-9}$). Next, we focused on the hazard rate for broken fixation after the movement cue was presented during the go cue delay duration. To this end, we separated trials by target cue value (Fig. 2.3 B), to estimate the probability of leaving the center port during the go cue delay. In this case, we show that briefly before any go cue could be given (before 250 ms, shaded area in the panel), animals have a higher probability of breaking fixation for rewarded trials. But, from that moment onwards, animals are more likely to leave the center port for non-rewarded movements.

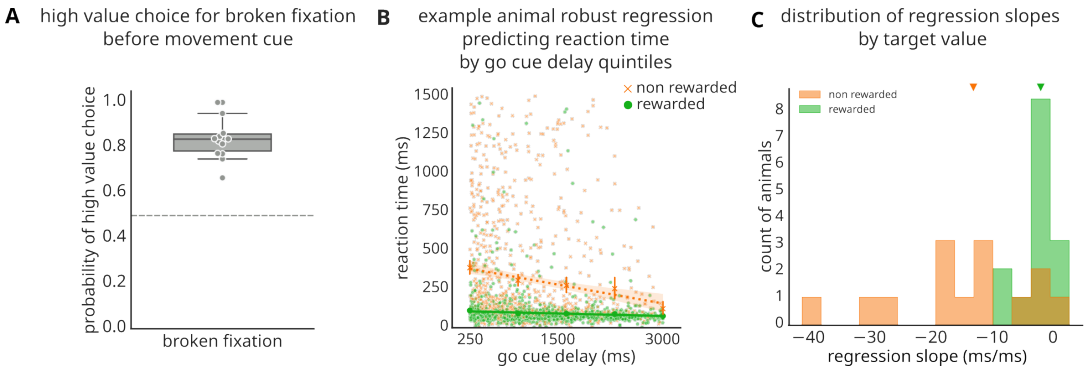


Figure 2.3: Broken fixation patterns and reaction times reflect the presence of a default motor plan. A. When breaking fixation before the movement cue, animals are more likely to go to the rewarded side. Probability of going to the rewarded port for broken fixations before the movement cue presentation, each dot represents the probability of going to the rewarded port for all the session for each individual animal. For broken fixation responses, we only consider responses made in the 1500 ms after breaking the fixation. **B.** Robust linear regression for all valid correct trials for a single example animal. Predicting the reaction times given the go cue delay duration and trial value. **C.** Animals take more time to initiate movements towards non rewarded location: Distribution of regression slopes from the regressions for every animal spliting rewarded and non-rewarded targets. The slopes of the animal in B are marked with lower triangles with colors marking their label. Panels B and C share the same color coding, *orange* for non-rewarded, and *green* rewarded target.

Another consequence of a default motor plan, is an effect of the go cue delay duration on RT that depends on the target value, and focus on the RT for valid and correct trials. If animals were to have a prepared motor plan, regardless of the go cue delay duration, they should have similar RT when they are sent to the rewarded side. Whereas, for non-rewarded targets, they would need to update their original plan, occupying some time. As a consequence, for shorter delays, animals would produce slower RTs and the longer the delay, the quicker we might expect them to respond, as they would have had more time to switch away from their default plan. To evaluate this, we fit, for each animal, two robust linear regressions, one for rewarded and the other for non-rewarded RT. We use as predictor the go cue delay

duration divided into quintiles and compare the resulting distribution of slopes and intercepts. We show the scatter plots of go cue delay duration and reaction times, colored by target values, and the resulting regression line for one example animal (Fig. 2.3 B). When comparing the distribution of the fitted predictor values for each regression line (Fig. 2.3 C), a Mann-Whitney test shows that non-rewarded trials have smaller slopes ($mean = -14.41$ms/ms, $SD = 11.87$ms/ms) than rewarded trials ($mean = -2.16$ms/ms, $SD = 3.06$ms/ms), $U = 28$, $p = .001$. This verifies that the slopes for rewarded trials are centered in 0, whilst the slopes for non-rewarded are mostly negative. These observations are consistent with our hypothesis: animals need additional time to update their original plan and settle in a new one, given the current trial requirements.

2.2.3 Behavioral signatures of inference as indicators of a task model

Our task comprises two possible contexts and context-driven behavioral responses also have another relevant implication: if animals were to learn that there are two discrete context, once they receive an unexpected outcome from a cue-response contingency, they could update the context in which they believe they are in. This would allow them to modify their responses globally, including responses for the non-experienced target. To evaluate whether this was the case, for each animal, we split transitions depending on the value of the first and second trials in the post-transition block, this gives 6 possible trial type groups. We split RTs for the groups into deciles and calculate the median value per decile, excluding the lower and higher ends, the range of deciles in (0.1, 0.9). We present the resulting data as a box and whiskers visualization in figure 2.4, where we separate trials by trial order after transition, the second trial in relation to the first, and value of the movement. In the figure, the two leftmost boxes are the data from the first trials after transition; the next two are the data for second trials when animals were sent to the non-experienced location; and the last two bars includes data for second trials towards the same location. We compare the distribution of RT deciles between groups, considering the animals as blocking variable. An omnibus test, Friedman rank sum test, indicates that the groups are different between themselves, $\chi^2(5) = 60.16, p = 1.12 * 10^{-11}$. To evaluate the differences between group, we perform pairwise comparisons using a Conover post-hoc test, using as blocking variable the animal identities, comparing between the groups. We use the Bonferroni correction to adjust p values for multiple comparisons. The result of this comparison is presented as the lines connecting bars in figure 2.4.

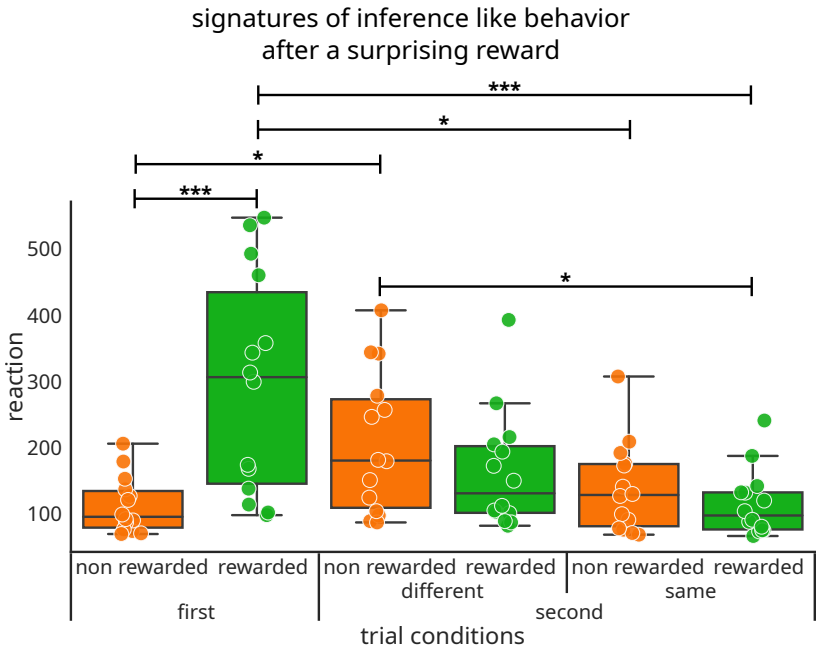


Figure 2.4: Rats display signatures of inference-like behavior after transitions. Box and whiskers plots of reaction times for the first two trials per block by target value, after conditioning blocks by first and second target values, in rows and columns, respectively. For box-plots, we aggregate all values across sessions, and dots are the medians over sessions by condition. In *orange*, non-rewarded targets; *green* rewarded. We label the panels with *same* or *different* depending on if the second trial is towards the same or to the other location with respect to the first trial respectively. To convey statistical significance, we use *** for $p < .001$; ** for $p < .01$; and * for $p < .05$.

The first difference is within the first trials, where RTs to both trial types are significantly different between each other, $p = 2.9 * 10^{-4}$. As this are the first trials after the transition, animals are unaware of the change in contingencies, and initiate movements to non rewarded location quickly ($mean = 124.06$ms, $SD = 68, 22$), and slower for rewarded ($mean = 349.75$ms, $SD = 284.85$). For the second trials, there is a significant difference between trial that followed a rewarding outcome ($p = 1.9 * 10^{-2}$), second non rewarded different ($mean = 273.73$ms, $SD = 253.61$) and second rewarded same ($mean = 130.03$ms, $SD = 79.78$). For trials after a surprising non rewarded outcome, there are no significant differences between the groups ($p = 4.9 * 10^{-1}$) of trials in second rewarded different ($mean = 203.60$ms, $SD = 159.14$) and second trial non rewarded same ($mean = 165.25$ms, $SD = 119.86$). These comparisons are indicative of an update after a surprising reward for the non-experienced outcome, but that this is not the case when the surprising outcome is non-rewarding. This is also noticeable in the differences between first trials rewarded and second trials to the same location ($p = 4.4 * 10^{-4}$), and first non-rewarded and second non-rewarded different ($p = 1.4 * 10^{-2}$). As both comparisons imply that animals changed their responses after experiencing a surprising reward as a first result after a transition. Interestingly, even though after a non-rewarding first trial animals do update their responses toward the non-experienced side, if sent a second time towards the same non-rewarded target, they have still not fully updated their RTs, we interpret this as if they were being optimistic in their outcome expectations.

In summary, these observations imply that after surprising outcomes, the animals are updating their behavior to both experienced and non-experienced targets. This represents an inference-like mechanism, as they have made a change in their responses not only to the experienced target, but also to the non-experienced one. But, they are not fully updating their responses in the case of surprising non-rewarded outcome, animals maintain fast RTs to the previously rewarded target after two consecutive non-rewarded outcomes. This over-optimism could appear, because it takes longer for them to update their responses when learning from non-positive outcomes where novel information is weighed differently depending on the difference in obtained value with respect to the expected (Sharot, 2011).

2.2.4 Embodied signatures of context-dependent default motor plan

One way in which animals could be preparing movements to a particular side, depending on their knowledge of the context, is to orient themselves in a way that could facilitate this movement. To check if animals are in fact embodying their preference,

we consider their pose at trial initiation. In case they were to have a default motor plan, they could enter trials orienting themselves to facilitate reaching the target where they expect to receive reward. In a single session from two animals (Fig. 2.5 A), we can see that their body position at trial initiation are more similar depending on reward location, as can be seen by the contrasts presented. The almost specular appearance of the images, implies that animals are orienting their bodies differently when they expect the reward to be on the left or right. This orientation difference is made clearer if we look at the average medial axis from all frames with respect to context (thick colored lines on the panels in Fig. 2.5 A). When looking at all individual animal across sessions (Fig. 2.5 AB), we observe the same general pattern. Animals initiate trials with a particular body orientation depending on the reward location for the current context.

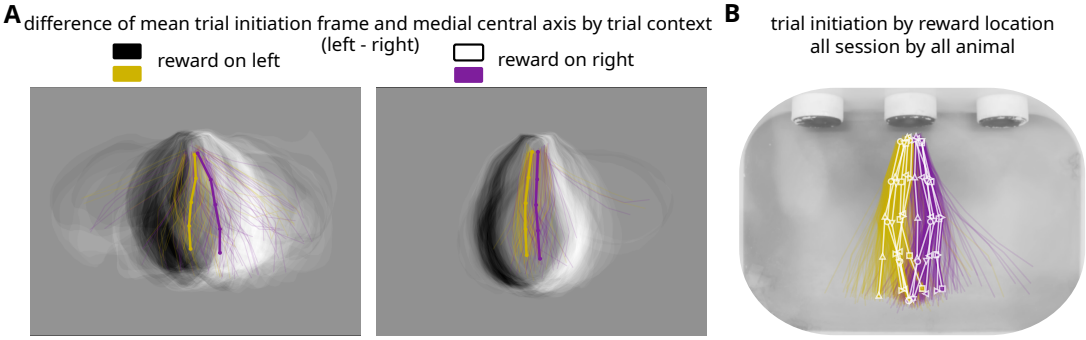


Figure 2.5: Animals initiate trials in with particular body orientations depending on context. A. Each panel shows the difference between mean first frame at trial initiation by context for a whole session for two single animals. Given the contrast, left — right, in *black* are the trials where reward would be given on the left, and in *white* when reward would have been given on the right. Overlaid to the contrast, we draw the central axis for each trial mean central axis of each individual frame colored by context, *mustard* reward on left, *purple* reward on right. Thinner lines are the individual frame medial axis, thicker lines their mean. **B.** Median central axis for first frame at trial initiation for all six animals, with video recordings separated by context. The background is the reference frame used in analyzes, ovelayed are the individual session for all animals, thin lines, and in white the median by context for each animal across sessions. Each animal is labeled by the marker in the white lines. The colors for the lines are the same as in panel A.

2.3 Discussion

The presented results imply that in the presented delayed movement task, animals learn the presence of two discrete contexts. This is noticeable in the observations that animals prefer rewarded over non-rewarded trials and that they initiate movements towards the targets with different RT distributions for rewarded and non-rewarded trials. On the one hand, if animals waited for the movement cue, they are more

likely to break fixation for non-rewarded trials during almost all possible go cue delay durations. Supporting the idea that animals have a preference for the rewarded trials, and are willing to take a time out to avoid going to the non-rewarded side. On the other, the RT distribution for non-rewarded trials appear to be scaled with respect to the distribution associated to rewarded trials, implying an active procedural cost in the production of these movements. Consequently, we pose that this effect is not only driven by motivational value, or vigor, but also by the presence of a default motor plan. A loaded program to go towards the rewarded location that has to be modified if the animal is cued to do the non-preferred action. We derive this interpretation from the behavioral patterns of broken fixations before movement cue, the distinctive evolution of RT over go cue delay duration by cued target value, and the differences in postures at trial initiation given contexts. For animal choices in broken fixation before movement cue presentation, animals are more likely to select the port where reward would have been given for that block, implying a bias for that location. The differential effects by target value over delay duration is noticeable in the manner in which go cue delay duration affects RT for trials that will be or not rewarded. A default motor plan to go to the rewarded side, implies that there should be negligible effect in the RT for the rewarded trials, whereas for non-rewarded trials it should take longer for the animal to initiate the movement. Because of this, for non-rewarded movement, short go cue delays have longer RT, and the longer the go cue delay, having more time to update their original plan, the quicker they are. The last evidence about the default motor plan is that this preference for rewarded location is embodied in the manners in which animals position themselves in the nose port at trial initiation, before being told where they will be sent. We also showed that knowledge about the context allows animals to build an internal model of the task, allowing them to update their responses to both experienced and non-experienced contingencies after a surprising outcome. As we showed that a surprising outcome in one of the targets, is followed by a change in how they respond to the other. There seems to be a difference in the effect of this update when the surprise is positive or negative: positive surprising outcomes drive a faster update of the behavior. Importantly, we are able to notice this default motor plan given our manipulation on the delays and the freely moving nature of our task, as they give the behavior of the rats more space to express their intentions.

In conclusion, we provide evidence that animals build an understanding of the different context in the task; that they are not only responding with a difference driven by vigor, but also by preparing a movement towards a preferred location in each context; that they are updating this plan, depending on trial contingencies; that they embody this knowledge about context in how they initiate trial; and finally, that

they are using knowledge about context to update their responses globally after experiencing a surprising outcome. The similarity of our results with previous studies in primates support the presence of a conserved mechanism between mammalians, which allows the preparation and update of motor plans depending on expected outcomes. These observations highlight the degree of flexibility of the cognitive mapping of the animal model organism, that allows them to map contexts and state of the environment into generalizations yielding inference-like behaviors. These forms of biasing action selection, by means of learning and updating mappings from previous experience, is a hallmark of the reinforcement-learning framework. This operationalization of agent-environment relations, allows modeling and exploring how an agent learns about how his actions drive environmental changes, from which the agent receives different rewards. These contingencies, allows the agent to learn mappings of states, actions and outcomes, and to select better policies given the expected outcomes of the current state and future ones.

The cortico-basal ganglia-thalamo-cortical circuit has the relevant architecture and connections to implement motor controlling and behavior biasing processes given actions-outcomes expectations, to prepare a default motor plan, and update the action plan once the task demands it. Taking into account the connectivity of the circuit, we expect that output regions of basal ganglia (BG), such as substantia nigra pars reticulata, should send relevant signals to motor regions of the thalamus, ventral anterior and lateral nuclei of the thalamus. As the nigrothalamic pathway is the shortest path that BG reinforcement-learning like algorithms can use to inform or bias cortical activity towards the selection of preferred responses and update these programs to changes in environmental demands. In the next chapter, we present the results from recordings of neurons in the output of the BG and motor thalamus, and propose how they relate to the observed results.

2.4 Methods

2.4.1 Animals

A total of 14 adult male Long-Evans rats between ages 5–16 months were used in this study, all acquired either from the Champalimaud Foundation Vivarium or Charles Rivers Laboratories. All animals were housed in groups of two to four animals per cage, under a regular 12 h dark/light cycle, with lights ON cycle starting at 8:00 am. After the first day of behavioral training, they were weighted and maintained under water-deprivation for the rest of the experiments with ad-libitum access to

food. Their weight and general well-being was assessed daily before experiments, and their participation in experiments was stopped if weight decreased more than 80% of their original baseline. In which case, they were single housed and given a week of ad-libitum access to water, their weight after this period was considered their new baseline and experiments resumed.

2.4.2 Behavioral apparatus

All experiments were conducted in a behavioral plastic box (TROFAST, IKEA), with dimensions 36 cm tall, 22.5 cm wide and 35 cm long. The box interior had three equidistantly distributed nose ports at floor level in one wall, one speaker, and was lighted from a custom-made lid that allowed video recordings. The nose ports were 3d printed and housed a white LED. An infrared emitter-sensor pair that allowed the detection of port entries and exits, and their outside facing layer allowed the accommodation of a printed circuit board (Champalimaud Foundation Scientific Hardware Platform). The two lateral ports were also equipped with a metallic spout connected to a 20 mL syringe via a solenoid valve (LHDA1231215H, Lee Company). A microcontroller board (Arduino Mega 2560, Arduino) was used to monitor and control all the sensors, peripherals, and actuators through a finite state machine. The produced and detected port events and other task related data were timestamped and serially communicated to a computer desktop with a Windows 10 operating system and stored in a text file using a python script.

2.4.3 Behavioral assay

2.4.3.1 Pre-training procedure

Animals were gently situated inside the box 5 days a week for sessions that lasted 2 hours. Before the experimental task was introduced, they were trained to relate the box elements and their responses to task variables. The procedure started by acclimatizing them to the box and to learn the reward delivery condition, for one session both lateral ports LEDs are lit and any entry in them results in the delivery of 25 µL of plain water and an auditory tone (1750 Hz, 150 ms). The next day, we start training them to learn the trial based task structure and the contingencies between cues, ports interactions, and reward availability. Every trial after the first, starts 9 s after the previous, this is the inter trial onset interval (ITOI). Before starting a new trial, one of the lateral ports is randomly selected, and trial availability is signaled when the center port LED is lit, and it starts with center port entry. This leads to a

brief white noise (1 ms), the put out of the center LED, and the selected lateral port LED is lit, this is followed by an auditory go cue (7000 Hz, 125 ms). Entry in the lit port results in reward delivery, and entries to the unlit one are inconsequential. From this session onwards, animals are trained to withhold movements while maintaining their snouts inside (fixating) in the center port for longer periods of time. To this end, in every trial after selecting the lateral port, we randomly sample a duration from a Gaussian distribution that increases in steps of 10 ms in mean as animals perform correctly, and decreases 2 ms if they make a mistake. The sampled duration is used as the delay between the center poke entry, and the auditory go cue. Every trial where animals sustained the center port fixation for the required time, a lit side port entry is rewarded, but entries in the other ports are inconsequential. Leaving the center port prematurely for more than 30 ms, breaking fixation, leads to the side port LED to be unlit, a burst of white noise (150 ms) and a time-out. That implies that the next trial will take 10 more seconds to start. We repeat this training procedure until animals are capable of performing correctly for more than 100 trials with durations sampled from a distribution centered at 4 seconds for 2 consecutive days. All tested animals took at most 3 weeks to reach this stage.

2.4.3.2 Task training procedure

After animals have learned the relevant contingencies between their actions, box events, cues, and the reward location, we introduce the possibility of making errors and a new delay. The procedure is split in two stages, the first stage consists in one session where each trial start by sampling two delays from a truncated exponential distribution centered at 1500 ms, with a minimum of 250 ms and maximum of 3000 ms. If a sample were to fall outside this range, we sample another until both values lie within these boundaries. The values are assigned to a movement cue delay and a go cue delay. Afterward, a side port is randomly chosen and the center port LED is lit. Following center port fixation, this LED is put out, and if the animal maintains the fixation for at least the movement cue delay, the selected side port LED is lit. If the animal sustains their snout in the center port for the go cue delay duration, the auditory go cue is given and a lit side port entry is rewarded. If the animal were to enter the unlit lateral port, a white noise burst (120 ms) will be played, and a time-out of 10 s would be added to the ITOI. Breaking center port fixation before the go cue is given is also discouraged by a time-out. In the next stage, we introduce the animals to the case where not all side port entries will deliver reward. During this stage, all contingencies and trial structure are maintained, but every session, a

reward-delivery probability is chosen. During this session, after delays and side port selection, a random value is sampled uniformly in the range [0, 1], if the value is above the fixed reward-delivery probability. In case the animal were to maintain fixation in the center port for the duration of the sum of the two delays, entry in the lit side port will not result in reward delivery. Animals stay in these stage until they are capable of obtaining more than 7 mL of water during two consecutive sessions, with a reward-delivery probability of 0.75. Animals fulfilled this training requirements in a week, and no animal took more than 2 weeks to achieve this performance.

2.4.3.3 Task procedure

After animals had learned all the aforementioned cue and behavior contingencies, they are introduced to the delayed movement task. At the beginning of each 2 h session, one of the side ports is randomly chosen as the rewarded side, and a block duration value is sampled uniformly within the range [30, 40]. Before each trial, we sample a 'movement cue' and a 'go cue' delay from a truncated exponential distribution in the range [250, 3000]ms, with mean 1500 ms, also a target side port is pseudo randomly selected. Trials after the first one are available with an ITOI of 9 seconds, that is, each trial starts after at least this time has elapsed since the initiation of the last. Trial availability is signaled by the LED in the center port being lit. Once the animal fixates his snout inside the port, the LED is put out, a brief 1ms white noise tone is played, and a stopwatch starts running for the duration of the movement cue delay. If the animal maintains the center fixation for this period, the target side port LED is lit to signal the target direction, and a second stopwatch start running for the duration of the go cue delay. In case the animal were to keep the fixation in the center port for this delay, a brief auditory tone is played (7000 Hz, 150 ms) as a go cue signal, and he can leave the center port. Once the animal leaves the center port, marking the reaction time. If he reaches the lit lateral port, the LED is unlit, another brief tone is played (1750 Hz, 150 ms) to signal the correct choice, and the trial would count as valid and correct. If rewarded side and target side were to be the same, 0.25 µL water reward is given, in other case, there is no reward would given. In the event the animal were to leave the center port before the go cue tone plays, a brief white noise would be played (120 ms). The trial, would be counted as a broken fixation, and a time-out would be given to the animal (adding 10 s to the ITOI). Similarly, if the animal were to fixate up to the go cue, but entered the unlit port, the same error white noise tone would be played and a time-out given, but this would count as a valid error trial. After the block-duration number of trials, the rewarded side is flipped

without any signals about this given to the animal, another block duration is sampled and the session continues. The first four trials after a transition, target directions are sampled uniformly random. After these, all target directions are subject to a pseudo-random selection procedure. To this end, we keep track of the left and right report probabilities, regardless of being correct or error, and sampling a random value for selecting a left direction trial. We also estimate the probabilities of responding to the rewarded or non-rewarded trial, and draw a random value of selecting a rewarded trial. We then pick the maximum between probabilities for laterality and value and consider them the respective biases. If the laterality bias is higher, we compare the probability of selecting a left direction trial with the leftward reports probability, if the random value is higher, we choose the left target direction, else we choose the right. If the value bias were higher, we compare the probability of selecting a rewarded trial with the probability of making a rewarded choice, if the first is higher we sample a rewarded trial, else we sample a non-rewarded. In case of a tie between the laterality and value biases, we use the sampled choice left and choice rewarded direction values as the biases. This procedure enforced that animals had to make the same number of valid trials for both rewarded condition and target locations, controlling for possible value or laterality biases. We also use a correction-loop procedure during the task, if the animal were to make 3 incorrect responses for the same trial type, we would only give them those types of trials until the number of errors dropped below 3. Correction-loop trials are not counted as valid, and their responses have no effect on the counter of trials in the block.

Behavioral analyses

For all animals, we removed from all analyses the first 20 sessions after being introduced to the task, as during this period their behavior is still crystallizing. We also remove from each animal sessions where their performance for valid trials was below 80%; session with more than 30% broken fixations of the total number of trials; sessions in which they did not respond for more than 5% of the valid trials; for each session we calculate a directional bias as the total number of trials to the left over the total number of trials for valid and correct, and for broken fixation. We further removed session with a directional bias outside the [30%, 70%] range in any condition. Finally, except when noted, we also removed the 5 trials after a transition and the first block of each session. To estimate the linear regressions in figure 2.3 C, we use the Theil-Sen estimator, implemented in SciPy (`scipy.org`). For the multiple comparison analyses, we used the scikit-posthocs library (`scikit-posthocs.readthedocs.io/`).

To estimate the hazard rate H of breaking center fixation in the discretized time interval $\hat{T} = \{k_1, \ldots, k_n\}$, with $T = t + \Delta t$, where $t \in \mathbb{R}_+$, we use the following equation:

$$H(k_i) = \frac{B(k_i)}{\sum_{j=k_i+1}^{n} B(j) + C(j)}$$

Where $B(k_i)$ and $C(k_i)$ are the counts of broken fixations and completed trials at the i^{th} interval, respectively. Thus, $\sum_{j=k_i+1}^{n} B(j) + C(j)$ is the number of broken fixations and completed trials occurring after the interval up to the longest cue delay in $\hat{T}$—3 s.

2.4.4 Video acquisition and analysis

A digital camera (Flea3 FL3-U3-13S2, Point Grey Research Inc.) was mounted outside and over the behavioral box. It acquired and recorded the video in the same desktop computer that was used to save the behavioral data. The stream was acquired using a custom Bonsai workflow (Lopes et al., 2015) and saved them at 30 FPS in 1280x960 pixels in 8-bit grayscale resolution. The camera received a TTL pulse at every trial start from the microcontroller to ensure alignment of task events with the video frames. To get the location of the animals in each video, we trained an artificial neural network model to do image segmentation, for this we used the segmentation models python library (Iakubovskii, 2019)(`https://github.com/qubvel/segmentation_models.pytorch`. Briefly, we sampled 210 frames from different sessions, manually labeled the images with masks to mark the locations of the rat and the 3 nose ports. Then we sample 147 of the frames (70%) to be used as training set, 48 to use as validation, and the remaining 15 are used as test. We instantiated a UNet++ architecture (Zhou et al., 2018) with an EfficientNet encoder with 17 M parameter pre-trained in the 2012 ILSVRC ImageNet dataset, and trained it to segment the animals. The F1-score was used as a loss function, and we tracked the Jaccard index as an accuracy metric in the validation set. For training, we augmented our data using PyTorch albumentations library (`albumentations.ai`), we used the horizontal or vertical flip with probability of 0.5; shifted, rotated and or scaled the images; cropped a 320x320 pixel square and added Gaussian noise to the sample, we varied the image perspective with probability of 0.5; applied one of two possible transforms, either Contrast Limited Adaptive Histogram Equalization (CLAHE) or randomized the image gamma; choose one of the following transformations sharpen, blur or add motion blur to the image; changed randomly the brightness and contrast, or random-

ize the HSV values of the image; and finally, we did one of three possible non-rigid transformations, either an elastic transform, a grid distortion or an optical distortion. These augmentation procedures were applied to both the cropped image and mask. This scheme allowed us to make efficient use of our relative small dataset. The network is trained for 300 iterations, and the resulting model was capable of segment images from the validation dataset with high accuracy (Jaccard index of 0.95 in the test dataset) as shown in figure 2.6 panel A. We trained another similar model to segment the nose ports from the dataset (after training Jaccard index 0.9 in the test dataset, example result in figure 2.6 panel B). We use this second model to align videos from different sessions, we took a still image of one of the boxes to be the general reference frame, and for each session we randomly sample 200 frames and take their median. From these general and session reference frames, we segmented the location of the nose ports, and then calculated a rigid affine transform that would put the ports in the single session in the location of the general reference. This procedure allows us to put all recorded session in similar a coordinates frame, thus making distances between them comparable, an example of this procedure is depicted in figure 2.7. To analyse single trials, we find the central axis of the mask in the frame. For this, we take the mask of warped single frames, remove any masked areas with less than 10K pixel^2 and find the central axis of the region by means of a skeletonization operation. To transform the pixels into a line, we find the location of the pixels in the medial axis, find the two edge points, and sort them by their Euclidean distance starting from the point closer to the central port. After sorting, we interpolate this medial axis line to 100 points and keep 5 equally spaced points, starting from the one closer to the central port and ending in the last end point.

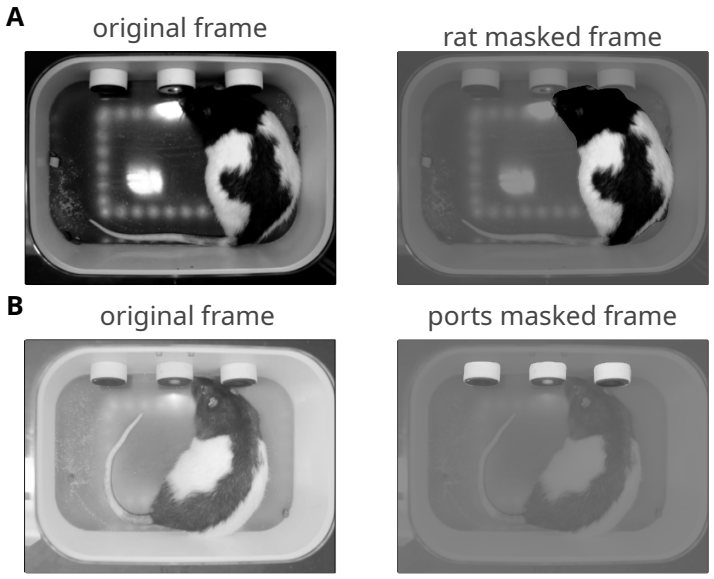


Figure 2.6: Segmentation of the rat and nose ports in video frames. A. The image shows an example frame from the test set after training the segmentation model to mask the animal. In the top row are the original frame and mask, on the bottom the predicted mask and the overlay between the mask and frame. **B.** We also trained a segmentation model to mark the pokes in the boxes. Top row shows an example frame from the test set and the mask for the poke associated to it, bottom row show the predicted mask and the overlay of the mask and frame.

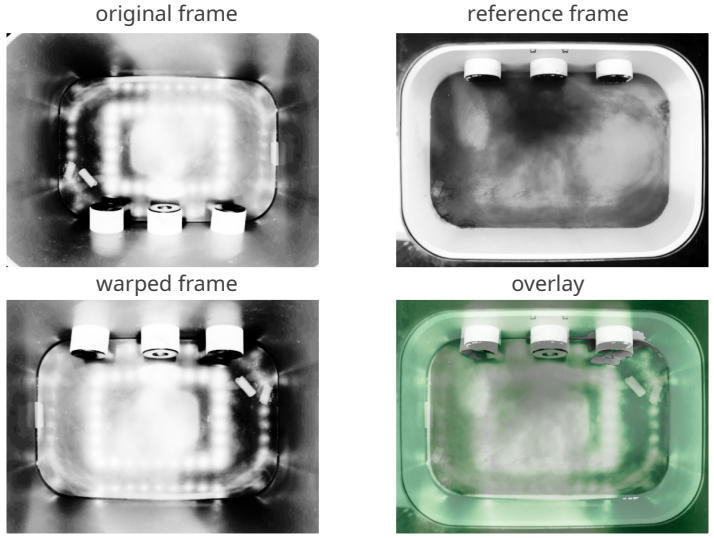


Figure 2.7: Warping of video frames with respect to the poke locations. Using the port locations from a reference frame, we compute a rigid affine transform to warp a frame into the same orientation and general spatial distribution. The top row depicts the original and reference frame, the bottom shows the warped original frame and the overlay of the warped and reference frames.

Chapter 3

Basal ganglia output and thalamic correlates of context dependent preparation

> And now, for something completely different.
>
> Monty Python

3.1 Introduction

Recognizing the underlying setting or circumstances that give meaning to an event —what in general language we would call a context— gives animals the advantage of being able to prepare the actions they expect to be more beneficial. The ability of grouping states of the world and being able to understand that similar actions might imply different outcomes, allows animals to anticipate how to face changing situations, once they understand how they relate to previously experienced ones. In the previous chapter, we observed how, through their behavior, animals understand that our task has two distinctive contexts and initiate trials with a plan to go towards the rewarded port. This default motor plan is updated when the target location is different to the desired, and implies a cost in the time it takes them to initiate the movements towards the non preferred target. From a modelling perspective, animals can be considered agents confronted with an environment over which they have to act, and from which they receive observations and rewards. This is an operational definition that frames the problems from a reinforcement-learning perspective. From this view, the animal needs to learn in which way the available actions that the environment can receive will lead to better cumulative outcomes. To allow this, the agent needs to perceive environmental cues, relate the actions taken and their outcomes to the underlying

states of the environment, and learn through experience what policies would better serve for the long term goal. Current interpretations of the cortico-basal ganglia-thalamo-cortical (CBGTC) circuit indicate that this circuits contains all the relevant and necessary elements to perform these computations. Cortical activity, representing the environment, actions, and contexts, is transmitted to the basal ganglia nuclei (BG). Here, dopaminergic (DA) signals modulate synaptic weights at cortical and thalamic target regions in the striatum, allowing for selection of actions that have resulted in better outcomes from the presented possibilities. These mappings learned over experiences, will drive the activity of the two synergetic pathways internal to the BG, that will communicate to output regions which actions to enforce or suppress depending on the expected outcomes.

A large portion of the cortical mantle receives sensory stimuli from the internal and external environment, and organizes it in hierarchically and structured components (Bennett and Hacker, 2012; Kandel, 2013). This organization allows the development of maps, e.g., body or retinal (Welker, 1971; Tootell et al., 1988) that can facilitate the organization of what is available and possible to be done (Cisek, 2007). As an illustration, primary and premotor cortices have been shown to produce signals related to deliberation and commitment during decision-making (Thura and Cisek, 2014); and signals in the frontal-eye-field in primates and frontal-orienting-field in rodents show motor selection signatures before saccades or orienting movements (Hauser et al., 2018; Erlich et al., 2011; Boyd-Meredith et al., 2022). These cortical contents can be considered a state representation, that can interact with motor and expectations signals to facilitate responses to environmental calls for action. Higher in the cortical hierarchy, the frontal lobe, known as a higher order associative region (Nauta, 1972), has been shown to encode for these expectation signals (Schweimer and Hauber, 2005; Oswal et al., 2007) and being relevant in reward-guided learning and decision-making (Rushworth et al., 2011). Below the cortical mantle, the basal ganglia (BG) input area, striatum, receives organized projections from the many cortical and thalamic regions (Hunnicutt et al., 2016; Hintiryan et al., 2016), and sends modulatory signals that will excite or inhibit motor plans via BG output regions (Mink, 1996). The cortical mappings are also maintained in striatum, and seems to be a relevant feature present along the different nodes of the BG (Romanelli et al., 2005; Nambu, 2011). Dopaminergic (DA) nuclei intrinsic to the BG, the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNc), shape the synaptic weights at corticostriatal and thalmostriatal projections (Centonze et al., 2001; Shen et al., 2008; Gerfen and Surmeier, 2011) facilitating the maintenance of a policy-like mapping of states and valuable actions. The interplay of striatum and DA signals will also affect other

two intrinsic regions of the BG, the subthalamic (STN) and the external segment of the globus pallidus (GPe). All share the structured connectivity present in both cortex and striatum van Dijk et al. (2016); Iwamuro et al. (2017), and selectively modify the activity of the output regions of the BG, namely substantia nigra pars reticulata (SNr) and the internal segment of the globus pallidus (GPi), the homologous in rodents being the entopeduncular nucleus (EP). Striatal inhibitory projections, and the interactions of GPe and STN inhibitory and excitatory projections, respectively, shape the direct and indirect pathways that increase or decrease the inhibitory outputs regions (Gerfen and Bolam, 2010). The interplay between excitatory and inhibitory projections seems to be at the base of precise motor control (Chen et al., 2021), and the production of goal-oriented behaviors, by suppressing competing actions (Cruz et al., 2022). BG output nuclei, GPi and SNr, have shown relevant motor signals, as can be expected (Benhamou and Cohen, 2014), and also convey value signals (Yasuda and Hikosaka, 2015) to their downstream targets. These output structures project to motor controlling regions in the brainstem and cerebellum, but also send projections to motor related nuclei in the thalamus (Sakai et al., 1998; Nishimura et al., 1997; Kuramoto et al., 2011). These motor thalamic (MTh) nuclei also send projections to cortex and striatum, showing a closed loop architecture (Parent and Hazrati, 1995a; Kuramoto et al., 2009; Foster et al., 2021). As already mentioned, along the processing pipeline, there seems to be a parallel and hierarchical structure maintained, where cortical territories and hierarchies are conserved in the receiving structures (Kim and Hikosaka, 2015; Hooks et al., 2018; Foster et al., 2021; Maurin et al., 1999). This has lead to the notion that the CBGTC circuitry could be sustaining multiple parallel representations that access only partial observations of the environment at hand (Lau et al., 2017). The original oculomotor 1DR-ADR task (Kawagoe et al., 1998) and subsequent modification, have greatly informed about the roles of BG in shaping the reward orienting responses of animals (Hikosaka et al., 2000, 2006, 2014). And, the interplay between the CBGTC network are good candidates to be the source of changes in corticospinal excitability associated to expected rewards in effectors before action execution (Klein et al., 2012; Bundt et al., 2016), which could be driving behavioral effects such as the ones we previously demonstrated in the previous chapter (see chapter 2).

Taking into account our task relevant manipulations on value based movement preparation and reward expectation (Fig. 3.1 A) the roles of CBGTC nodes in the learning and computations of these relevant features, and the architectural constraints of the circuit. We decide to focus on the BG output and MTh (Fig. 3.1 B), as this link is the shortest path that BG computations can take to reach cortex (Gerfen and

Bolam, 2010). From the outputs of BG we focus on SNr as it has the largest cell population (Oorschot, 1996), receives denser inputs than GPi (Foster et al., 2021), and has been shown to carry reward expectation signals (Bryden et al., 2011). From the MTh, we focus on the ventral anterior and lateral (VA/VL) nuclei, as it has been shown to receive from SNr and mainly project to motor cortices and striatum (Sakai et al., 1998; Kuramoto et al., 2009; McFarland and Haber, 2000, 2002). We focus on the activity around trial initiation, movement cue presentation and port-out event as this are key moments within the task, where the animal has access to different levels of information. During trial initiation, the animal is only aware about the context of the current trial; at movement cue presentation, the animal is informed about the target direction and, with this information, can also be aware of the value of the upcoming movement; finally, before port-out, the animal already is committed or has updated a motor plan that is going to be initiated. During these events and in regions of interest, we expect that relevant task features should be detectable in BG output regions and in motor thalamus. Given our task features or dimensions, we expect to find signatures of: context of the current trial, direction of the upcoming movement, expected value of the movement or interactions between them (Fig 3.1 C). Given the structure of task contingencies, and these circuit nodes relevancy for the communication of actions and assigned value expectations to them, we expect both regions to have signatures of the aforementioned features. Importantly, value signatures should be observable after context and direction are acknowledged, as this feature depends on the interactions of the other two. Finally, by taking into account the modulatory roles of the main cell populations in both regions over their downstream targets, we expect differences in the stability of the signals during the period after go cue and movement initiation. In this sense, as BG output is mainly inhibitory of downstream areas, it should maintain the inhibition of a selected plan, regardless of value; whereas, given MTh excitatory role on cortical and striatal populations, it should mainly facilitate the updating of a new plan given task contingencies. In particular, during shorter delays, the stability should change quickly to facilitate the update of the default motor plan. We first show that animals in these experiments have similar behavioral results to non-implanted ones, then we focus on single neuron correlates of task events, and finally we move to population analyses by means of linear decoders. We use linear decoders in two different manners: First, we use them to show how decodable information about context, direction, and value evolves over time. Secondly, we use them to evaluate the stability of the information in the population activity between the initiation of the selected movement and once the target has been reached.

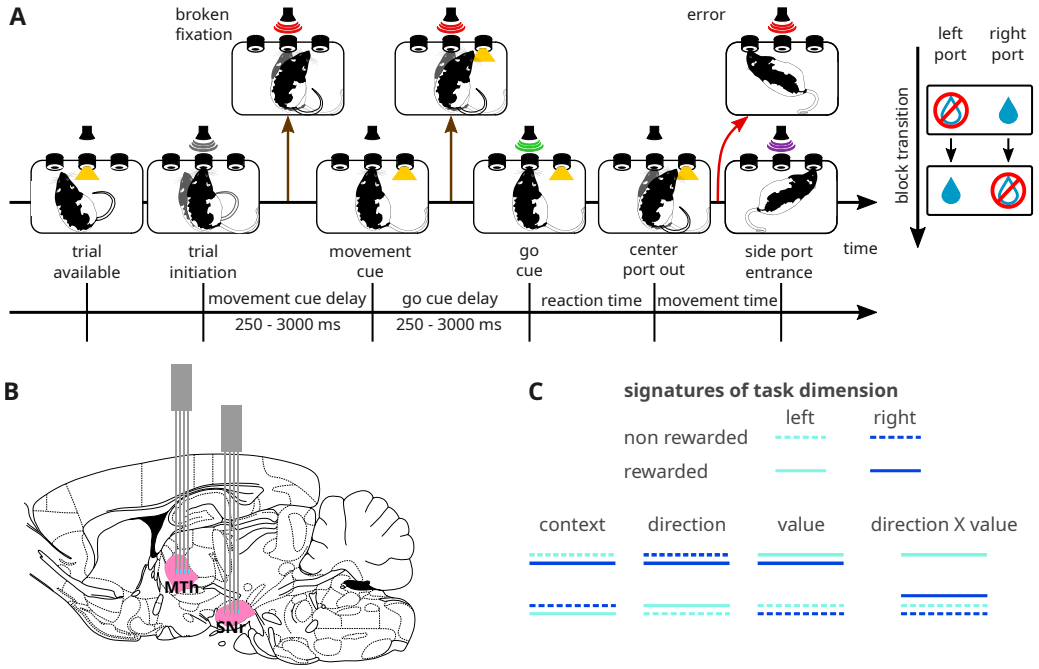


Figure 3.1: Task description, implant location and hypothetical signatures. A. Task description, we used the same procedures as in chapter 2. Briefly, animals are required to initiate trials in a center port, and wait for two delays whilst maintaining the fixation. After the first delay, one of the lateral ports LED is lit, and after the second delay an auditory cue is given to inform them that they can leave the center to go to the cued location. Importantly, in blocks of trials the animal will only receive rewards in one of the lateral ports, and this contingency will change without informing the animal. If they were to leave the center port before the auditory cue or go to the non cued location, a time-out would be given, for an in dept explanation the reader is invited to look at section 2.4.3.3. **B.** Location and arrangements of the electrodes in the target regions, we target motor nuclei of the thalamus (VA/VL), and substantia nigra pars reticulata (SNr), surgery procedure and implant coordinates are described in section 3.4.3. Location od target region are highlighted in pink. Image adapted from Paxinos and Watson (1998). **C.** Color coding to be used when depicting single cell activity, and hypothetical task dimension signatures that could be present in single cells. Non rewarded trials have dotted lines, and rewarded continuous; to mark target direction we use *cyan* for left targets and *blue* for right, ipsiversive and contraversive, respectively. For task dimensions, a pairing of different colors and line styles implies contextual information; pairing of line colors implies direction; pairing of line style value; and, isolation of one line style and color indicates an interaction between direction and value.

3.2 Results

To ensure that the procedure did not affect negatively the performance or relevant behavioral signatures, we evaluate whether implanted animal show context dependent action mapping, and a default motor plan that updates by task demands. We analyze behavioral responses using the same procedures described in Chap. 2. We first evaluate if rats display context dependent mapping of responses. In the stable trials, we observe that RTs are affected target value; implanted rats are slower for non rewarded trials in comparison to rewarded (Fig. 3.2 A). In consonance with non-implanted animals, RTs for non rewarded targets are scaled with respect to the rewarded (Fig. 3.2 C). In sum, these observations indicate that implanted animals are aware of the context of the block in which they are in, and that they initiate their responses driven by the expected outcomes.

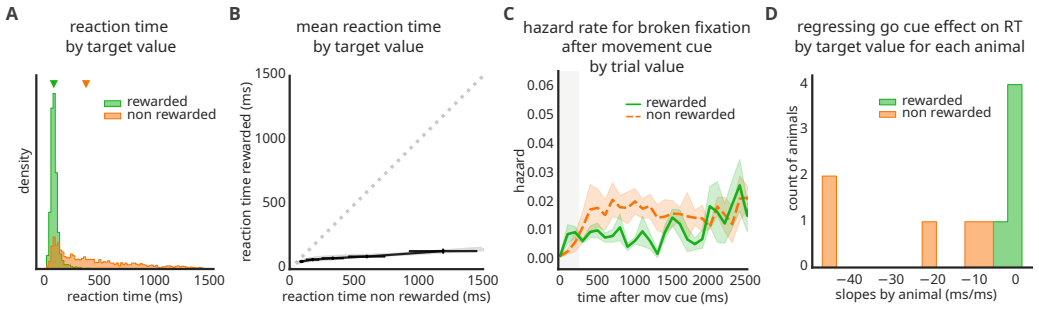


Figure 3.2: Implanted animals display all relevant behavioral correlates of context understanding and motor plan preparation. **A.** Marginal distribution for all valid trials with respect to target value. The colors used are *green* for rewarded, and *orange* for non-rewarded. **B.** Median reaction times for all sessions by animal divided in quantiles, marginalized by target value. We use as abscissa the value for non-rewarded and ordinate the rewarded, In *black* is the mean over animals, and *gray* each individual animal. **C.** Hazard rates for broken fixations after movement cue presentation over go cue delay duration by target values. We calculate the individual hazard rates by animal and present the mean and SEM, *orange* for upcoming non-rewarded movements and *green* for rewarded. **D.** Distribution of robust linear regression coefficients for the 5 recorded animal, the model was fitted to predict RT from the go cue delay duration. In *orange* coefficients for non rewarded, *green* rewarded.

When evaluating if the implanted animals also display the behavioral signatures associated with a default motor plan that they update by task demands (Fig. 3.2 C & D). We observe that, on one hand, for broken fixations after the movement cue presentation, they are more likely to abort trials for non rewarded targets (Fig. 3.2 C). When we fit robust linear regressions to predict the reaction time using the go cue delay duration as predictor, and compare the distribution of coefficients. We notice that coefficients for non rewarded movements are in general negative, whilst those of rewarded are all centered at 0 (Fig. 3.2 D). These results imply that implanted rats

are willing to wait longer for rewarded movements, but for non rewarded they prefer not to. And, that they initiate the trial with a plan to towards the rewarded location, but that they update this original plan with a cost in reaction times.

In sum, these observations are indicative that implanted rats are expressing the same behavioral signatures as the non-implanted. This allows us to start looking for the relevant correlates of these behavioral patterns in our electrophysiological recordings.

3.2.1 Single neuron in both BG and MTh show multiplexed correlates of task dimensions and behavior

The first step to relate the activity of the recorded population to our task is to observe single cells responses around the task events in different trial types. After organizing trials by context, movement direction and value, we can average the number of spikes of the recorded units in time bins around relevant task events, to build their peri-stimulus time histograms (PETH). We show one representative units from each recorded region, to highlight how the information carried by the single units varies over time and task events, indicating a kind of multiplexing of information. Where, a single neuron, or channel of information, carries different signals at different time points.

In one representative single unit SU recorded in SNr (Fig. 3.3), we observe a higher density of activity shortly after trial initiation for trials when the reward would be given on the left, as noticeable in the raster plot and PETH. After movement cue presentation, the number of spikes increases for trials where the animal will be sent to the right and receive a reward. And finally, around the initiation of the response, port-out, there is a higher firing rate in general for movements to the right. Thus, we can say that this single neuron response is changing during trial events; first informing about the block type, or context, that the trial is in; after movement cue the response shows an interaction between direction and value, and finally when the animal is going to initiate the movement towards the target, it informs about the direction of the movement.

The example neuron recorded in MTh (Fig. 3.4) also shows the context signal at trial initiation. But, after movement cue presentation, the main increase in activity seems to be related to the value of the upcoming movements, and this is sustained until after the animal initiates his movement. We also notice that these unit varies

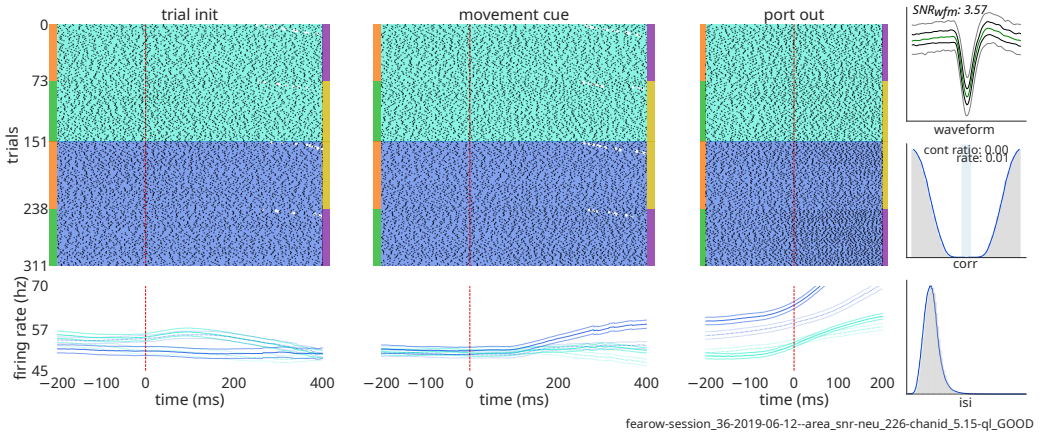


Figure 3.3: SNr unit shows different activity at behavioral events. Single cell raster plots (top row) and PETHs (bottom row) for the behavioral event, ordered in the three first columns. On top of each column, we add the descriptive signature that is observed from the analysis. Colors at the edges of the raster plots help to group trial types: *orange* for non rewarded trials and *green* rewarded; *mustard* and *purple* mark trials in blocks where reward would be available on the left or right, respectively; *Cyan* and *blue* group trial where target movement was to the left or right, respectively. PETHs color and line styles reflect trial type, continuous lines are trials that will be rewarded, dotted trials without reward. Lines in *cyan* are averages for trials with target direction to the left, and *blue* to the right. In both raster and PETH, we include a vertical *red* dashed line to mark the onset of the event. For raster plots, we add a white dot to mark the occurrence of the following task event. The panels in the rightmost column show the average waveform and its 50% and 95% CI, top; the autocorrelogram, middle; and inter-spike-intervals distribution (ISI), bottom.

his response over the trial events, carrying different information to downstream populations.

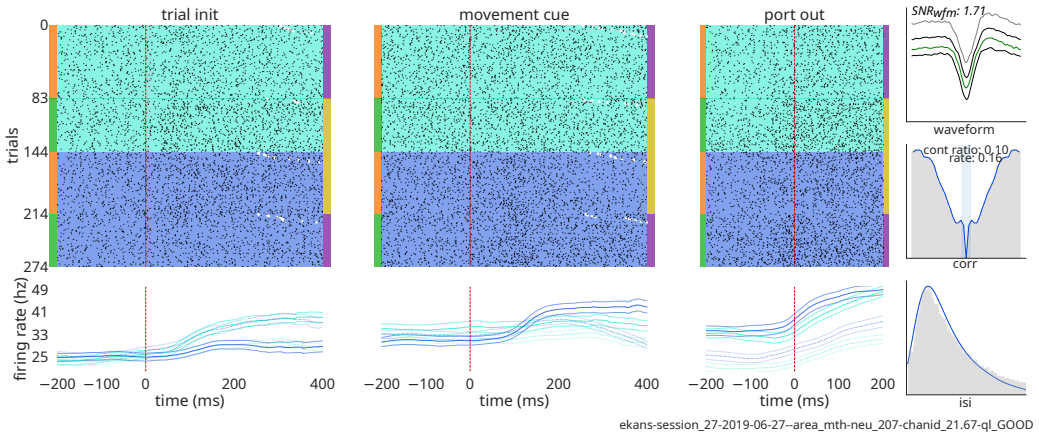


Figure 3.4: MTh unit shows different activity at behavioral events. Raster plots and PETHs from a single neuron in MTh during the task. The structure, color coding, line style and marks are similar to those in figure 3.3.

As we are interested in the reliability of the information that these neurons carry over time, we calculate the area under the receiver operating characteristic AUC-ROC curve. This is a single value that quantifies the ability of a binary classifier to discriminate between classes when the threshold is varied. In the single neuron case, the classes are defined by the trial class, and the classifier values are the spike counts of the neuron for each class. Briefly, this metric is calculated by integrating over the ratios between true positives and false positives when varying a threshold applied to the classifier responses. In this case, the threshold is the number of spikes per time bin for each trial type. The perfect score of 1 implies that the classifier would never err to predict the target class, and a score of 0.5 means that the classifier can not differentiate between the two classes. It is important to make note that a score below 0.5 means that the classifier is better at predicting the non-target class. To evaluate the performance of the AUC-ROC metric, we perform a permutation test. In the figures, we color values significantly different from the distribution of values with randomized labels (for a more complete description of the data processing and statistics, the reader is invited to look at this chapter methods in section 3.4.4).

For the well isolated SUs in SNr, we observe that AUC-ROC scores for context dimension (Fig. 3.5 top) are significantly high before trial initiation, different neurons carry information about current reward location, for both contexts. This information is sustained during movement cue presentation and at port-out, there is a higher number of neurons from whom we could read out in trials where reward would have

been given in the port contralateral to the recording site, the right. At trial initiation, there is no clear encoding of trial direction (Fig. 3.5 middle), as expected given that this information is not available yet to the animals. But, shortly after movement cue presentation, we find neurons that reliably inform when the animal will be asked to move to the port ipsi- or contralateral to the recording site, the left, or right respectively. At port-out, this information is also present and there are more cells are encoding ipsiversive than contraversive movements. Finally, when evaluating the target value of the trial (Fig. 3.5 bottom), we also see no decodability at trial initiation. After movement cue presentation, a large fraction of the cells are informative about rewarded trials. And, this information and relative presence is maintained around port-out. There is a broad literature on the effects of expected outcomes in the activity of SNr (Bryden et al., 2011; Sato and Hikosaka, 2002) and even during consummatory behavior by means of upstream D1/D2 MSN in striatum (Chen et al., 2021).

For the population of well isolated putative neurons in MTh, we start by looking at the information about the context (Fig. 3.6 A) and observe that it is possible to predict context at trial initiation. Around movement cue, the number of neurons from whom this dimension labels are decodable is reduced, and by port-out the decodability is still present but with lower values. For the direction of the upcoming movement (Fig. 3.6 B), we see no clear patterns around trial initiation, after movement cue presentation there are cells from whom one could decode ipsi- and contra-versive movements with respect to the recorded hemisphere. At port-out, there are also cells encoding also both directions. Finally, when predicting the trial upcoming value (Fig. 3.6 C), at trial initiation we do not observe neurons reliably encoding these classes. But, after movement cue presentation, from the large majority of cells we could be informed about rewarded targets movements, which is sustained at port-out.

To gain further insights about the kind of information reliably readable from well isolated single units independently in both regions. For each unit of time, we calculate the percentages of cells with AUC-ROC value significantly different from a shuffled distribution (Fig. 3.7[1]). We first focus on the SNr at trial initiation (Fig. 3.7 top row left panel), here notice that almost 15% of the population is reliably encoding context. Before movement cue presentation (Fig. 3.7 top center), the proportion of cells that encode context increases to about 25% in SNr and after presentation goes back to similar levels as before. Still after movement cue presentation, around 200 ms post cue, the proportion of independent cells reliable encoding movement and direction rise in tandem to about 15%. These percentages are maintained during port out for SNr (Fig. 3.7 top right). With respect to the number of cells encoding the different dimension, or combinations of them, reliably across time in SNr (Fig. 3.7

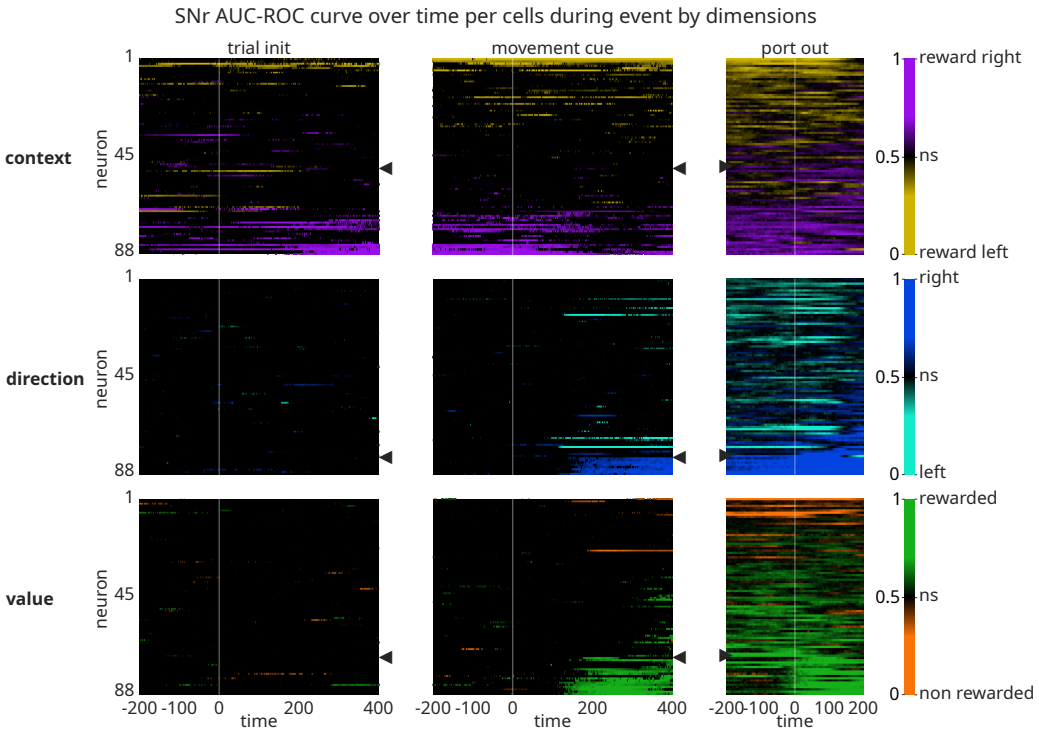


Figure 3.5: Single cell AUC-ROC scores indicate encoding of multiple task dimension by cells in SNr. All analysis were performed using only correct and stable trials during the task. We only color in the AUC-ROC p-values that were significantly different after a permutation test (for an in depth description of the method the reader is invited to look at the methods, section 3.4.4). **Top** Values for AUC-ROC in dimension context, we color with *mustard* significant decoding of left rewarded and *purple* significant right reward context. **Middle** AUC-ROC values predicting target direction, we color *cyan* significant decoding of leftward, and in *blue* significant rightward movements. **Bottom** AUC-ROC for target value, we use *orange* to color significant decoding of non rewarded target movement and *green* significant rewarded. Black triangles by the sides of the panels denote the location of the neuron presented as an example in Fig. 3.3. Neurons are sorted across events, by the norm of the vector of the maximum values of the AUC-ROC for each event. We mark with a *white* line the onset of the event in each panel. Color bars on the rightmost location indicate the label assigned to the values and color codes in the graphics.

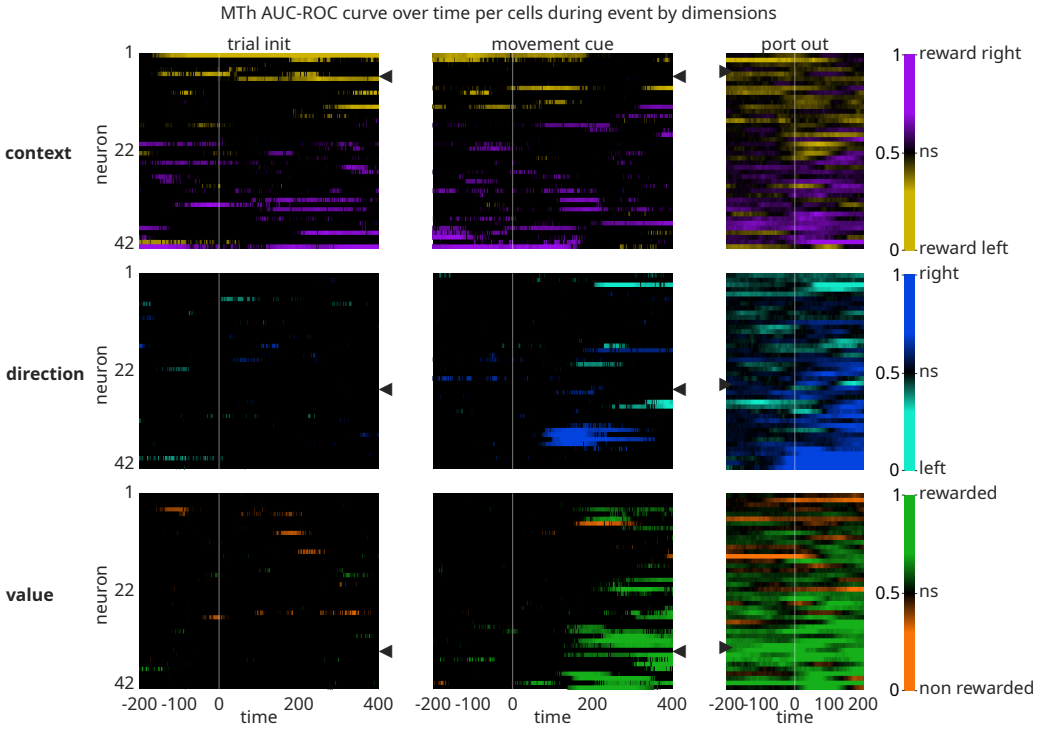


Figure 3.6: Single cell AUC-ROC scores indicate encoding of multiple task dimension by cells in MTh. Analyses follow the same structure and conventions as in fig. 3.5. **Top** AUC-ROC values when prediction of the trial context from SU. **Middle** AUC-ROC values predicting target direction. **Bottom** AUC-ROC for target value. Black triangles denote the location of the neuron presented as an example in Fig. 3.4.

insets in top row). We notice that at movement cue presentation, there are multiple cells encoding more than one dimension simultaneously. This latter observation is also present at port out, even though the majority of cells with high proportion of reliably decodable time steps per dimension are encoding value. We also find similar numbers of cells encoding context and value. In addition, at port out, we find that many cells are encoding more than one dimension. With respect to VA/VL or MTh at trial initiation (Fig. 3.7 bottom row left panel), we also observe nearly 12% of cells encoding context. After movement cue presentation (Fig. 3.7 bottom center), we observe an increase in the proportion of cells encoding upcoming movement before 200 ms have passed. The proportion of cells increasing in decodability for this dimension is in contrast with the decrease in proportion of cells encoding context around the same time event. The increase in proportion of direction encoding cells is followed by an increase in the proportion of neurons encoding value. Around the port out event (Fig. 3.7 bottom right), value encoding cells are the most prominent before movement initiation, and once the movement is about to start, the proportion of cells encoding direction and value increase. When looking at the number of cells in the MTh that encode the dimensions reliably for large portions of the event (Fig. 3.7 insets in bottom row), we also observe cells encoding more than one dimension. We see that at port out, the majority of cells that consistently encode task features are encoding value.

Our single-cell analyses give supporting evidence that individual cells within SNr and MTh are encoding the relevant task dimension during the task, and that in both populations there are neurons encoding multiple signal at different times. We observe that at trial initiation, both regions include a proportion of cells coding for context. We also notice that at movement cue presentation, neurons in SNr are consistently encoding information, whereas in MTh, the changes are more transient. Another difference in the activity of the single cells between regions is that SNr changes seem to evolve collectively, changes in one dimension are generally followed by changes in another. At port out, this collective behavior of SNr is maintained, and at this event, MTh also seems to follow the same pattern. As the proportion of cells encoding value increases, so does the number of cells encoding direction, even though there is still a difference in the relative proportion of cells encoding each dimension. If we focus on the number of units encoding reliably the different dimensions over time, we notice that both regions maintain relative similar proportions of cells encoding the

[1]The RGB values of the mean of the colors assigned to each of the two classes per dimension were transformed into hexadecimal, and the name of the most similar one in `icolorpalette.com` was used. Hex values are #AA6753, #0098D0, and #84932C; for context, direction, and value respectively.

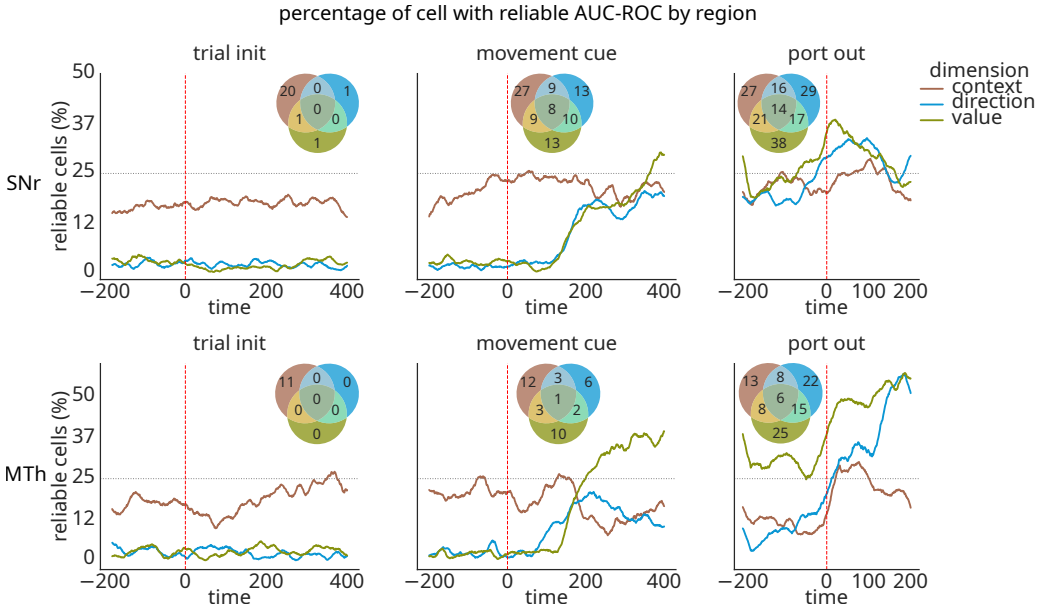


Figure 3.7: The independent cells in each region show dynamical encoding of the task dimensions. In the figure, we present the percentage of cells in each population that had AUC-ROC scores significantly different to the shuffled data at each time point. For the analyses, we included 88 SU from SNr and 42 SU from VA/VL thalamus. Each column of panels indicates one behavioral event, and each column is one population. In columns the events are trial initiation, movement cue presentation, and port out. In the rows, the recorded regions, on top SNr, on the bottom VA/VL nuclei of the thalamus. Each panel includes as inset the counts of cells that had more than 25% of points significantly different to the shuffled data during the event duration. The insets are Venn diagrams of cell counts per dimension and include the intersections between dimensions, if cells encoded more than one dimension during that period. In each panel, we color context with *SoHo red*; direction with *tomb blue*; and, value with *siskin sprout*. The colors of the insets use the same scheme as the task dimensions, and intersections of sets use the sum of the colors associated to each dimension. In the panels, we mark the moment of the event with a vertical *red* line, and add a horizontal *grey* line to mark the 25% of the population.

different dimensions and their combinations. This last observation points to a kind of multiplexing happening in these regions, where the same neurons are sending multiple signals at different times.

3.2.2 Population activity in both areas show differences in the temporal profile of information

Decoding task dimensions over time allows uncovering the information available to the population and how this process unravels. To this end, we use a cross-validated Monte-Carlo sampling method. In particular, we take a linear decoding approach using response vectors sampled from the whole population of putative neurons recorded. Briefly, we use support vector machines (SVMs) to find hyperplanes separating population responses into two classes. To avoid over-fitting, we train and test the algorithm in disjoint sets of trials (for a complete description: section 3.4.5). With this method, we evaluate the accuracy of our population to distinguish between the different task dimensions over time. This procedure gives us a way to access the information that downstream regions could decode from our recorded activity (Fig. 3.8).

After training the linear decoders with SNr population activity, our analyses indicate that trial context is decodable with higher than chance accuracy 125 ms before the trial initiation (Fig. 3.8 A left panel). This accuracy stays at a sustained high level for all other events. In contrast, as expected given the available information to the animals, only after movement cue presentation the accuracy for upcoming movement direction and value starts increasing (Fig. 3.8 A, middle panel). The accuracy for direction starts being decodable significantly better than chance 150 ms after the cue, decoding of direction starts being reliable at 190 ms after the cue. At port-out, the SNr population seems to reliably encode all relevant task dimension, nonetheless, direction reaches the highest accuracy (Fig. 3.8 A, right panel). These signatures of activity imply that SNr is informing downstream regions about context throughout trials, and that, after movement cue presentation, populations receiving these messages could be aware of the upcoming movement direction and value, in that order. Finally, by the time the animal will unfold the selected motor plan, direction is the main signal that this region is conveying to receiving populations, nonetheless, all dimensions are decodable accurately during this event.

When we evaluate the accuracy of the linear decoders over time on the MTh population, we observe that at trial initiation context is reliably decodable 20 ms before center port fixation (Fig. 3.8 B, left panel). After movement cue presentation, the decoding accuracy for context starts at 115 ms, followed by movement direction at

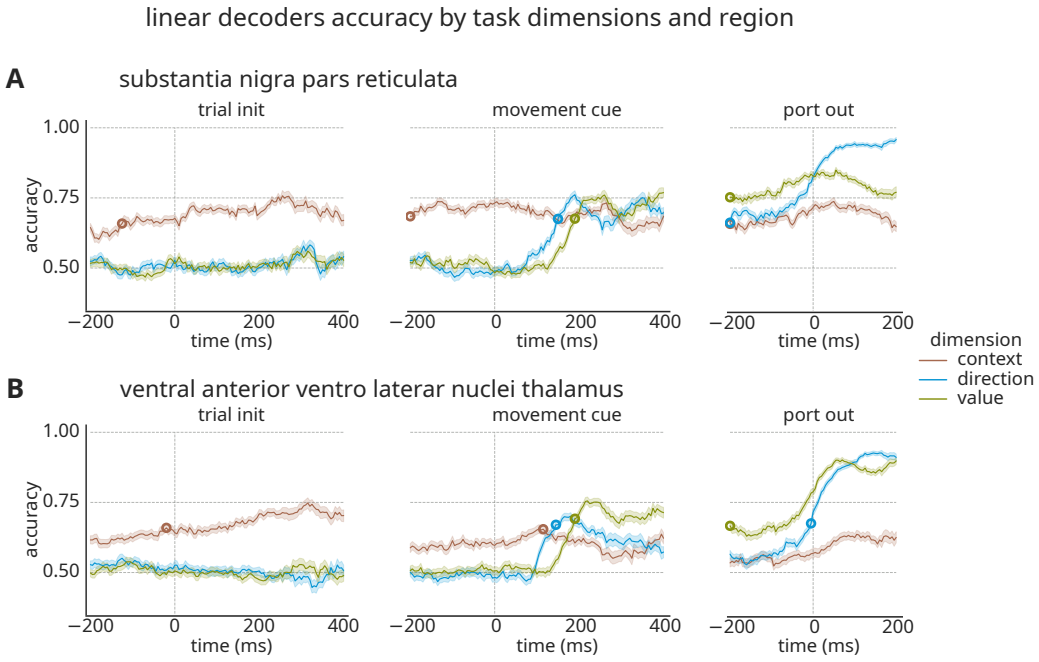


Figure 3.8: Relevant task features signals are reliably decodable in SNr and MTh. **A.** With the activity of SNr we can reliably decode context, and at port-out direction is the best decodable dimension, for the analyses we included 88 SU and 58 MUA from SNr recordings. **B.** From activity in putative neuron in MTh we can decode context at trial initiation, and at port-out direction and value, for the analyses we included 42 SU and 82 MUA from MTh recordings. Each column depicts the mean accuracy by dimension, and error bars are the 99% CI. We use the same line colors as in Fig. 3.7. The colored dots mark the moment when the accuracy of the decoder starts being significantly different from the shuffled distribution. If the dot is at the beginning of the line, the dimension is decodable from earlier than the minimum time; if the dot is absent, the decodability is not significantly accurate.

145 ms, and finally value at 190 ms (Fig. 3.8 B, middle panel). Still at movement cue presentation, the decodability accuracy for value stays in high levels, whereas, context and direction decodability decreases. Around the port-out event, accuracy for value initiates at a high level and is decodable earlier than 500 before movement initiation (Fig. 3.8 B, right panel). During the same event, just as the animals are about to initiate their action, 5 ms before initiation, decodability for direction increases. At port out, context decodability is never accurately decodable above chance levels. Taking this activity in consideration, it seems that the populations receiving these signals, would be informed transiently about the trial context; that the movement cue drives an increase in the decodability of the value of the upcoming trial; and, that at port-out, value is the most relevant signal that this region is communicating.

From the decoders over time, we notice a pattern in the time course of the decodability of the signals. On the one hand, at movement cue presentation, both regions seem to first encode direction and after a short delay value starts to increase in decodablility (Fig 3.8 A & B, middle panels). On the other hand, at port-out, SNr leads in decodability for direction and value in comparison to MTh (Fig 3.8 A & B, right panels), but value is decodable to a higher degree in MTh population activity.

3.2.3 Stability of decoders performance reveals distinct roles of the recorded regions

One way to assess the stability of the information available in a population is to use the so-called "temporal generalization", or "off-diagonal decoding", strategies. Here, multivariate pattern analysis algorithms are fit on one time point, to later be evaluated in others (King and Dehaene, 2014). As the linear decoder fits coefficients to optimally classify observed responses in recorded neurons, if the activity is stable, the classifier accuracy should remain constant over time. Whereas, if the activity changes into a different state, the decoder would be prone to errors. We are interested in the evolution of the responses after animals have been informed about the target location, during the go-cue delay, and how it relates to the activity after the execution of the motor plan, after the port-out. As we saw in the behavioral results, during the go-cue delay, animals need to update or commit to their default plan, and non rewarded trials RTs have a negative correlation with delay duration. Given this, we fit linear classifiers with the activity of the populations in a window of time after port-out, but before the animal has reached their target, focusing on the direction and value dimensions. To later evaluate the accuracy of the trained classifiers in predicting the label for responses before the go cue is given, splitting the delays in tertiles of the go

cue durations (for an in depth description, see methods section 3.4.6). In this way, the accuracy of our decoding scheme informs about the stability in dynamics from our recorded populations during these periods, and how these dynamics evolves during the time that animals have for committing-to or updating their motor plan.

For SNr activity, when evaluating decoding stability for direction (Fig 3.9 top), our analyses show higher than chance decodability for ipsilateral movements throughout the delay duration. Contralateral movements stay below chance levels, although shorter delays for rewarded trials start higher than chance and drop to chance levels afterward (Fig 3.9 top leftmost panel). With respect to the value dimension (Fig. 3.9 bottom), we observe a trend to stay at the threshold level of accuracy for all trial types. In particular, shorter delays for non rewarded movements quickly rise to the threshold levels (Fig. 3.9 bottom leftmost and center left panels). This patters of stability, imply that SNr has a stable encoding of ipsiversive movements, and that contraversive ones are associated with a reorganization of the activity during the preparatory period. They are also indicative that value information is somewhat stably encoded in SNr, and that for non rewarded trials it goes through a quick reorganization during the shortest delays.

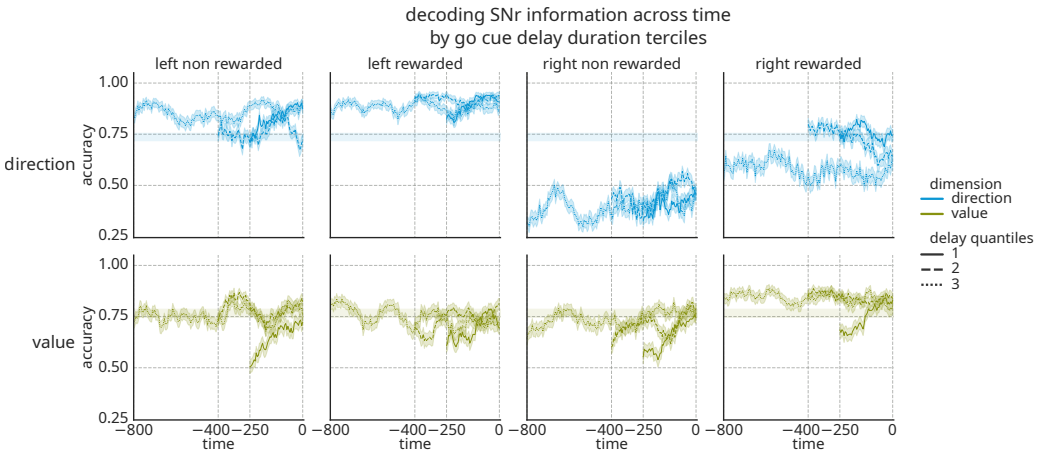


Figure 3.9: SNr presents a high level of stability of decodability for trials with ipsilateral movements with respect to the recording hemisphere, and overall high decodability of trial value. For the analyses, we included 88 SU and 58 MUA from SNr recordings. Top row depicts decodability stability for task dimension direction, bottom row for value. Vertical dashed gay lines indicate the median duration of the median and shortest go cue delay durations, 400 and 250 ms, respectively. Horizontal dashed gray lines indicate different levels of performance. Colored lines represent the mean accuracy per dimension, and their shadings depict the 99% CI. Horizontal bars represent the median accuracy of significantly accurate time point for the SVM decoders in figure 3.8 and a 10% margin around as confidence. Line style separates the different go cue delay durations, shortest for continuous; medium in dashed; and, pointed the longest. Colors are the same as in Fig. 3.8.

In the case of MTh, the stability for target direction stays at chance across all delay durations and trial types (Fig. 3.10 top). Whilst, decodability of value for non rewarded trials reaches higher than chance values, though higher for contraversive trials; and, for the shortest cue delay durations there is a ramping in the stability up to a plateau (Fig. 3.10 bottom). In sum, this pattern implies that MTh population activity necessary to decode direction is varying during the go cue delay. Whereas, activity that could be useful to decode the value of the upcoming trial is more stable for non rewarded trials, disregarding the direction of the movement to be performed.

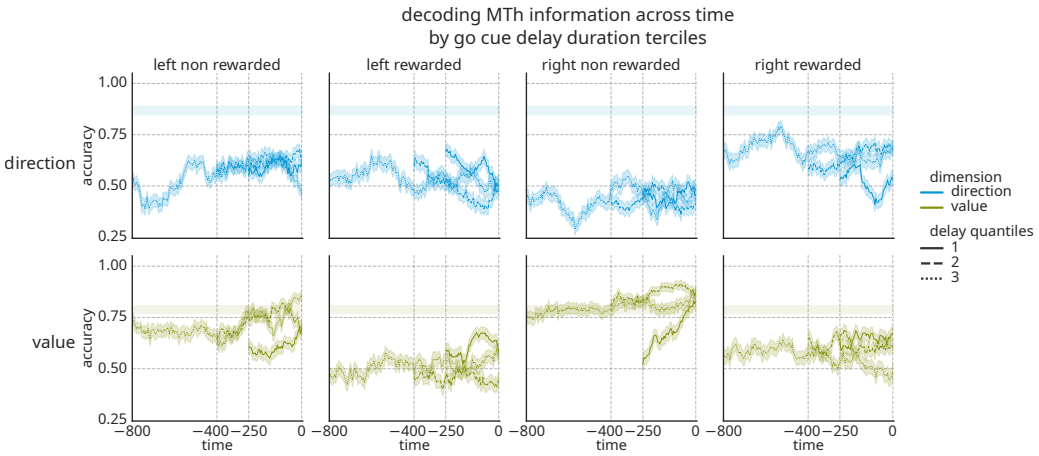


Figure 3.10: Stability of MTh is not stable for direction, but for value is higher for non rewarded trials. For the analyses, we included 42 SU and 82 MUA from MTh recordings. Panels are ordered, and use the same indications and color scheme as in Fig. 3.9.

3.3 Discussion

From the analyses of single cells, we notice that both SNr and MTh have neurons encoding the task contexts throughout the trials. This type of context relevant signals in SNr have been reported in oculomotor tasks in non-human primates (Handel and Glimcher, 2000). This type of signal have also been show in MTh and discussed as stable value signals (Yasuda and Hikosaka, 2018) that could drive PFC activity supportive of decision-making (Yang et al., 2022). We also show that after movement cue presentation and up to port-out, there is a dominance of neurons encoding for ipsiversive movements in SNr, whilst both regions seem to have a majority of neurons informative about rewarded trials. This is in line with previous studies showing motor and reward related signatures and projection in SNr in both primates and rodents (Hikosaka et al., 2006; Rizzi and Tan, 2019; Basso et al., 2005; Sato and Hikosaka, 2002; McElvain et al., 2021); the over representation of direction in SNr seems to be in

line with the area known projections and roles over motor controlling regions (Schultz, 1986; Kha et al., 2001; McElvain et al., 2021), and it has been shown that there are neurons representing ipsi- and contraversive movements in the region (Basso et al., 2005; Lintz and Felsen, 2016), and also in MTh (Catanese and Jaeger, 2021; Kuramoto et al., 2009). Plus the relevant, and well-organized projections that both regions share (Klockgether et al., 1986; Kha et al., 2001; Antal et al., 2014; Kuramoto et al., 2011; Gulcebi et al., 2012). The single cell analyses, also indicates than in both regions individual neurons are encoding not only one task relevant dimension, one neuron can encode one dimension at one time, and later a different one. As if neurons in these regions were multiplexing information. This has been proposed as one of the possible roles of SNr in the context of eye movements (Basso and Sommer, 2011), and also observed in midline thalamic neurons in Pavlovian conditioning (Li et al., 2016). Here we show evidence of individual neurons carrying different information which could be received by downstream populations to guide their local goals, for example to promote a particular action, or to signal a particular value expectation.

Population analyses over time, show that SNr encodes more consistently context throughout the trial, and by the time of port-out, direction seems to be the most relevant dimension. Whereas, MTh encodes context transiently during trial initiation, but at port-out, both direction and value are reliably encoded. After movement cue, both regions first encode direction and later value, and also both regions show encoding of value to a higher level of accuracy before port-out. It is noteworthy that SNr seems to be ahead of MTh in both direction and value around movement initiation, which agrees to their locations in the CBGTC. This evidence of contextual signals in SNr being more prevalent in comparison to the ones present in MTh, could be related to our recordings targeting VA/VL thalamus, known to project mainly to motor-regions. Whereas mediodorsal (MD) nucleus shares motor and prefrontal targets (Kuramoto et al., 2009, 2015; Bosch-Bouju et al., 2013; Çavdar et al., 2014; Xiao et al., 2009), but we still see above chance accuracy for value in both recorded regions.

The observed patterns of accuracy in the linear decoders over time, support the notion that BG outputs are leading MTh. This could be understood as if BG output projections are enforcing the update of the motor plan before the port-out event. Thus, shifting activity in cortical organization via MTh projections. This interpretation goes in line with the research that has in depth characterized the projectome of the CBGTC as parallel loops (Aoki et al., 2018; McElvain et al., 2021; Foster et al., 2021), and experimental evidence about the role of nigrothalamic projections in value coding, and behavioral update and control (Yasuda and Hikosaka, 2015, 2018; Hintzen et al., 2018; Inagaki et al., 2022).

Finally, in our analyses about the stability of the signals in both regions during the go cue delay period, with respect to the initiation of the action. We show that for the movement direction, ipsiversive trials have a higher stability in SNr, whereas contraversive ones seem to involve a reorganization of the activity in the region. For the MTh population, direction does not seem to be stably encoded during these period. For the value dimension, during the preparation of both rewarded and non rewarded movements the activity in SNr shows high stability, even for non rewarded trials the activity stabilizes quickly. The pattern of stability of MTh population indicates that rewarded movements imply a reorganization of the activity during the delay period, whereas, non rewarded movements, have a high stability. The patterns of stability for these regions, at population level during the go cue delay, can be understood in terms of the roles that these regions have in the CBGTC loop. Where SNr has a role in controlling the motor command to be released, thus is needed to drive the inhibition of the ipsilateral movements. But, is also informing downstream regions about the expected value of the upcoming target, explaining the smaller overall changes in stability for rewarded movements and the quick change for non rewarded. Whilst, MTh could facilitate a reorganization of the default plan. Given that non rewarded movements require an update of a prepared movement, cortical activity must be biased toward a new state, and this could be the role of the activity of MTh for this trial type.

In summary, the analyses of our recordings show that both regions contain a relevant context information, and that the interaction of this information with the movement information allows the rise of the value information on both regions. The results also show that at the time of action initiation, SNr presents earlier a direction signal, and that both regions encode expected value during this event. Even though previous experiments have shown that SNr leads value encoding with respect to MTh (Yasuda and Hikosaka, 2018), to our knowledge, we present the first evidence of time differences in the encoded information of activity at the population level in these regions for movement preparation. Moreover, our analyses of stability of the population responses over time support the notion that BG output is maintaining "the finger on the trigger" over the release of the selected action, and informing about outcome expectations; whereas, MTh seems to be facilitating modifications of cortical and striatal activity into new states, to update a previously selected plan.

3.4 Methods

3.4.1 Animals

A total of five adult male Long-Evans rats aged between 9 and 15 months were utilized in this study, procured from either the Champalimaud Foundation Vivarium or Charles Rivers Laboratories. Following surgery, the animals were individually housed under a standard 12-hour light-dark cycle, with lights on at 8:00 am. Post-surgery recovery, they were weighed and subjected to water deprivation for the duration of the experiments, while having ad-libitum access to food. Daily assessments of their weight and general well-being were conducted before each experimental session. Experimentation was halted if an individual's weight dropped by more than 80% of their original baseline or if their overall responsiveness was compromised. In such cases, they were provided with ad-libitum access to water, and their weight was monitored daily until recovery. Upon recovery, experimental procedures resumed.

3.4.2 Behavioral box, assay & analysis

The behavioral box and assay used for these experiments was similar to the ones presented in the previous chapter, the reader is directed to their respective section in the methods (Chap. 2, section 2.4), for behavioral box description see Chap. 2, section 2.4.2 and for task training and description Chap. 2, section 2.4.3. Importantly, we remove the correction loop procedure during the electrophysiological recording sessions. For behavioral analyses we also use the same procedures described in Chap. 2, section 2.4.3.3.

3.4.3 Chronic recording implant and data selection criteria

After achieving stable performance in the delayed movement task, five animals underwent surgery to receive two sets of electrodes targeted at the ventral anterior/ventral lateral (VA/VL) nuclei of the thalamus and the substantia nigra reticulata (SNr). The coordinates were centered with respect to Bregma at -1.9 mm medial-lateral (ML), -2.6 mm anterior-posterior (AP), and -6 mm dorsal-ventral (DV) for the VA/VL thalamus; and at -2 mm ML, -5.4 mm AP, and -8.6 mm DV for the SNr (Paxinos and Watson, 2009).Three animals were implanted with two independently movable silicon probes, each with either 32 or 64 channels and 2 or 4 shanks with 16 electrodes each (Cambridge Neurotech); one animal had two sets of single-wire

electrode bundles with 16 channels each in two cannulae (Innovative Neurophysiology); the last animal was implanted with a single-wire electrode bundle of 16 channels targeting the SNr and a 32-single-wire stiff electrode bundle implant targeting the thalamus (Tucker Davis Technologies). During surgery, the electrodes were slowly positioned 1 mm above their target locations and cemented to the skull with dental cement. Post-surgery, all animals received one dose of analgesics for 2-3 days (carprofen 5 mg/kg subcutaneously) to minimize discomfort, and they were provided with one week of ad-libitum access to food and water for recovery. Following this recovery period, animals were reintroduced to the task for 1-2 weeks, during which they were connected to the recording setup to acclimatize them to the procedure, but data were not saved. Post-session, the electrodes were lowered by 100-150 µm to reach their target locations. After surgery, two animals experienced issues with the implant in the medial thalamus (MTh), and no signal could be recorded from them.

After each successful recording session, we advanced the electrodes 100 µm to ensure that independent neural populations were recorded every session. Signals from the electrodes were amplified and digitized in the implants connected via a headstage (Intan Technologies). The electrophysiological signals and synchronization events from the behavior were and acquired at 30 kHz and 1kHz by a OpenEphys acquisition board and recorded by a custom-made workflow in Bonsai (Lopes, 2015) to a Windows 10 desktop computer. Recorded data was offline processed by means of a custom python program taking advantage of the Spikeinterface library (`https://github.com/SpikeInterface/spiketoolkit`), briefly data was bandpass filtered between 0.3-7.5 kHz, channels with too high noise removed, and common average referenced. Given that some animals movements could induce transient large fast changes in the signal, we removed such artifacts by detecting events with very large peaks (150 SEM) and setting a window of 10 ms around them to each electrode mean. Data from the single wire electrodes sorted and curated manually (Plexon offline sorter), and data from the silicone probes was automatically sorted by Kilosort 2 (`github.com/MouseLand/Kilosort`) and manually curated with Phy2 (`github.com/cortex-lab/phy`). After the curation procedure, we further label our putative neurons as multi unit activity (MUA) or single unit (SU). To do so, for each unit we calculate an average firing rate over the session; a waveform signal-to-noise ratio (Kelly et al., 2007); an inter-spike-interval contamination ratio (Hill et al., 2011); and, a presence ratio index, counting the number of minutes where the unit was present and divide by the total minutes for the session. To assign quality labels to units, we start by labeling all units as noise, then we evaluate each unit n in the whole dataset N, and update the labels to MUA

to those that satisfy the following conditions:

$$(FR(n) > 0.5\,\text{Hz}) \,\&(WF_{SNR}(n) \;>\; 1)\& \;(CR(n) \;<\; 0.4) \;\&\; (PI(n) > \;0.7)$$

Finally, we update the label to SU, to those n that satisfy the following conditions:

$$(FR(n) > 0.5\,\text{Hz}) \,\&(WF_{SNR}(n) \;>\; 1.5)\& \;(CR(n) \;<\; 0.1) \;\&\; (PI(n) > \;0.9)$$

Where $FR(n)$ is the average firing rate of n during the whole duration of the session; $WF_{SNR}(n)$ is the waveform signal-to-noise ratio of n; $CR(n)$ is the contamination ratio of m; and, $PI(n)$ is the presence ratio index for unit n.

We only used the responses for correct trials in all analyses. And, given the variable duration of the event, cues, and behavioral responses, unless otherwise noticed, for all event based electrophysiological analyses that use ranges of time around events, we remove spikes that would have occurred in a previous or subsequent event to avoid spurious information.

3.4.4 Quantification of single-cell response selectivity

For single cell response analyses, we use only the SU and remove non-stable trial from the recorded sessions. To assess response selectivity, we calculate the area under the receiver operating characteristic AUC-ROC curve for each task dimension independently. To achieve this, for each putative neuron, in the time range -1000 ms to 1000 ms centered at the event, we first make windows of 100 ms bins with 95 ms of overlap between windows. Then, sum the number of spikes in each bin, and use the result as the value for the last time stamp of the bin, to make this a causal analysis. We use the python library scikit-learn (`scikit-learn.org/`) to estimate the AUC-ROC metric for each time bin, using the trial types as labels. The algorithm takes the spike counts in each trial and calculate the rate between true positives and false positives that occur when using different numbers of spike counts as criterion. This way, one can build a graph that depicts the sensitivity of the performance of an unbiased observer using these criteria. A perfect classifier would never err, for any given criteria it would always assign the correct label. In such case, the area under the ROC curve would be 1. In case the classifier were to assign labels randomly, the area would be 0.5. And, importantly, a classifier that assigns the incorrect label to every target would have an accuracy of 0, a colloquial way of describing this case would be "*it is so bad, that is good*". We use the convention that for context, direction, and target

value dimensions, the lower bound of the AUC indicates reward in left, movement to the left, or non-rewarded; whereas, the upper bound is associated to reward in right, movement to the right, or rewarded, respectively. To calculate the significance of the AUC-ROC we applied a permutation test, (Efron and Tibshirani, 1993; Henderson, 2005). This implied that, for each neuron, we shuffled the labels of all trials and run the same analyses as before, and we repeat this process for 1000 iterations to get a distribution of possible AUC-ROC values. Afterward, we calculate the proportion of iterations that had a value equal or greater than the original to estimate an empirical p-value, using the method described in (North et al., 2002). To evaluate the significance of the AUC-ROC score, starting from an alpha $\alpha = 0.05$, we correct for the multiple comparison using the method from Bonferroni, desired $\alpha = \frac{\alpha}{NC}$ where NC is the number of comparisons. We considered significant only the values that were outside the $[{}^{\alpha}/_{2}, 1 - {}^{\alpha}/_{2}]$ range, a two-tailed test. In the visualization, we only color the values that were outside this range.

3.4.5 Population-level decoding analyses

For population analyses, we include SU and MUA, and we use all correct trial for every recorded session. For the population decoders, we first preprocessed the single neuron responses and then train a Support Vectors Machine decoder with a linear kernel, and regularization parameter $C = 1$ with the help of the (`scikit-learn.org/`) python library. Briefly, for all trials and events from each putative neuron, in the time range -1000 ms to 1000 ms centered at the event, we count the number of spikes in windows of 100 ms bins with 95 ms of overlap between them. The results are used as the value for the last time stamp of the bin, making this a causal analysis. We standardize this matrix of *number of trials* rows by *time steps by events* columns by subtracting to each value the mean and dividing by the standard deviation of all the samples. For the standardization procedure, we use all trials and events, but for all further analyses, we only kept the correct trials and the events presented in the figures. For each trial we also kept the reward location, cue delays durations tertiles, trial number in the block and other relevant trial labels. Given that the neurons were recorded from different animals and across many sessions, we made population 'response vectors' by sampling without replacement the activity of the neurons around the event for the same number of trials for each trial type. We were constrained by the minimum number of trials over all conditions, we used as trial type labels the movement and go cue delay tertiles, reward location, movement direction and movement value. These gave us a minimum of 6 trials per condition. For the training

dataset, we sampled 6 trials per condition per putative neuron from the longest cue delay tertiles, that is, for each event, we sample 6 trials from the longest upcoming or previous cue delay duration. For the evaluation set, we removed one trial per condition from the training dataset and also sampled 6 trials per condition from the left out cue delay duration tertiles. By sampling each trial by neuron events from different trial types, we remove all trial-by-trial correlation that could be present. Because both train and test samples are taken from different sets of trials quantiles, we are sure that we are not evaluating the performance of our classifiers in trials from the training sample, thus avoiding overfitting. With our train and evaluation datasets, we fit one SVM classifier by task dimension for each time-bin using the values and labels from our training set, thus generating three classifiers, for task context, direction, and value. After fitting them, we calculate the accuracy of each classifier in the evaluation dataset, comparing the predicted labels with the real ones. We repeated this procedure 250 times, to obtain a mean and confidence intervals of the accuracy of the classifiers. The binary linear SVM is defined as the linear operation: $f(x) = W^T X + b$, where W is a set of coefficient to apply to the observed spike counts per neuron, and b is a scalar threshold. After fitting the SVM, when evaluating an observation with this equation, it will be assigned a label of 1 to the values above the decision boundary, and 0 those below. The boundary is defined with respect to the set of observations that lie closer to the nearest edges of the clusters, *support vectors*. Those are the most difficult to classify, and the margin is the surface that lays between them. For a more in thorough description of SVM and their value in neuroscience the reader is invited to look at Koren (2021). To estimate the significance of the accuracy of the classifiers, we use a permutation test, (Efron and Tibshirani, 1993; Henderson, 2005). For this, we repeat the sampling procedures to generate a train and evaluation datasets, but shuffle the trial labels for each sampled response vector. Then we fit a classifier for each time-bin with the training dataset and evaluate in the test response vector the accuracy of our classifier. We repeat this procedure for 1000 iterations to get a distribution of accuracies of our bootstrapped samples. Finally, we calculate the proportion of shuffled iterations with accuracies greater or equal than the median accuracy of the non-shuffled classifiers to get an empirical p-value, using the method described in (North et al., 2002). To assess the significance of these result, we use $\alpha = 0.05$, i.e., if the proportion of values was greater than α, we consider it significantly different, one-sided test. We update our original α value to correct for pairwise family error using the Bonferroni method. In the figures, we used the 99% confidence interval from each classifier by task dimension to express the variability in accuracies over non shuffled

response vectors. We marked the first time that the decoders accuracies started being significantly different from the shuffled-labels distribution.

3.4.6 Stability of population information analyses

To estimate the stability of the information representation in the populations, we used the same data preprocessing and selection as in the previous analysis, but, we evaluated them using the ideas set by the temporal generalization method (King and Dehaene, 2014). In our use case, we trained SVM classifiers in the population response vector after the port-out and evaluated their performance in the response vectors before the go cue, calculating the accuracies by go cue delay duration and trial type. Briefly, for the training dataset, we sampled without replacement the population response vectors from the time range 0 ms to 200 ms after port-out, animals during these period never reached the target side port. And evaluated the accuracy of the classifiers in the population response vector from activity sampled in the range −800 ms to 0 ms with respect to the go cue. We repeat this procedure 250 iterations to derive a mean accuracy and confidence intervals. The classifiers achieved perfect classification on their respective training dataset, we also evaluated them by shuffling the training data session label 1000 times and computed their accuracy. In this shuffling procedure, the accuracy for all trial types for dimensions direction and value never fell below 75%, thus we are sure that they are a good fit for the training data.

To estimate the significance of these results, we calculate the median accuracy of the classifiers over time that were significantly more accurate than the shuffled (see section 3.4.5). We consider that if the decoders across time reach levels of accuracy above the median plus 5%, the population of decoders was predicting accurately the classes, and thus the population activity was stable. The "off-diagonal-decoding" procedure is relatively new, thus there are no clear ways to evaluate their performance. But, our proposal tries to accommodate for this issue by setting conservative thresholds that take into account the decodability of the information when using the available information.

3.4.7 Immunohistochemistry and microscopy

Histological analyses were performed to confirm the correct location of the recording electrodes. Rats were euthanized with a lethal dose of sodium pentobarbital (Eutasil, get dose for rats) and perfused transcardially with a solution of 4% paraformaldehyde. After skull extraction, brains were kept for 24 hours in a 4% paraformaldehyde

solution and later kept in PBS until sectioning. A vibrotome or criostat was used to section the brains into 40 µm to 50 µm thick coronal slices and stained by with NISSL to mark cellular nuclei. Later, images were acquired with a confocal microscope (LSM 710, ZEISS) or a slide scanner (Axio Scan.Z1,ZEISS). The location of the electrodes was validated afterward in the histological slices containing the target locations by following the lesions left by the electrodes or silicon probes.

Chapter 4

General Discussion

"You are decided, then, not to comply with my request——a request made according to common usage and common sense?(...) I would prefer not to."

Bartleby, the Scrivener,
Herman Melville

4.1 Overview of main results

Being prepared has tremendous advantages for animals that have to respond quickly to an ever-changing world. Being able to interpret the environmental cues and recognizing the context that gives meaning to the available actions, allows animals to anticipate what actions are more advantageous in particular states of the world. The main goal from our project was to understand the way in which the BG can modify cortical activity to guide adaptive actions selection. To achieve this, we first show how adaptive action selection unfolds when animals are demanded to make similar movements with different outcome expectations that vary after an unpredictable number of repetitions. Given this, our task allowed animals to prepare a preferred action, which we called default motor plan (DMP), but they also learned that sometimes they need to update this plan to get the most of each session. Importantly, in the context of our task, rats not only needed to update their immediate plan, to make the non-preferred action, but also the general mapping between actions and outcomes. As the context switches, the rats had to update their preferred action. In chapter 1, we showed that mammalian brains have structures and connections between them that facilitate be-

ing prepared, which allow learning flexible mappings for actions and outcomes. In this chapter, in section 4.2, we discuss the results from chapter 2. Where we show that we implemented a rat version of the 1-direction rewarded (1DR) task—delayed movement task, and characterized relevant behavioral effects of the manipulation in reward schedules. Next, in section 4.3, we discuss the results from chapter 3. Describing how analyses indicate correlates of the relevant task features at particular task events, and relate this signals to processes of commitment or update of the motor plan. On section 4.4 we present two proposals of synthesis of our results. We draft a sketch for a model that explains the relevant behavioral observations, that requires the relevant signatures observed in the data. We also include the preliminary results of an end-to-end model that captures elements of our data, and promises interesting future developments to put to test new hypothesis about the inner workings of the cortico-basal ganglia-thalamo-cortical (CBGTC) loop. Furthermore, we later discuss some limitations of our work and proposals for future steps in section 4.5. Finally, in section 4.6 we try to give an overarching closing statement of the relevancy of our work.

4.2 Behavioral signatures of expectation and context

As we described in chapter 2, animals trained in our delayed movement task recapitulated the results from primates in the 1DR task. By manipulating the delay of cue presentations, we gained further insights about movement preparation. The knowledge about the context, and their default motor plan, allowed the rats to present evidence of inference-like behavior. And finally, we also show that our freely moving version of the task allowed rats to express their knowledge about the context in the way they initiated the trails in each block.

In consonance with the results with primates, we first observed that animals are aware of the two possible outcomes based on the speed in which they initiate their movements after the go cue. Similarly to primates, rats initiated movements quickly for rewarded targets, and slower for non-rewarded. A second observation is that they have difficulty committing to the non preferred movement, as the scrivener in Melville's story "they prefer not to" go to the non-rewarded location. This effects can be related to a motor vigor effect, as rewarding movements would be promptly executed and non-rewarding are less enforced. Albeit, this can be a consequence of the DMP to go to the rewarded side. The first approach we took to evaluate if this was the case, was to look at what happened once animals abort trials. A DMP that is readily available

for them, implies that when animals are not able to wait fixating for the movement cue, they should be unable to withhold the execution of it. In line with this, we notice that animals are more likely to go towards the rewarded target if they abort a trial before receiving the movement cue. So, when a motor plan is initiated before the cue is given, they execute their DMP. Another consequence of a DMP to go to the rewarded side, is that rewarded movements should start quickly but non-rewarded must take longer to initiate. The longer the fore-period between movement cue and go cue, the quicker they can be. Having to do something different from what is loaded implies that animals need to update the action plan. As rewarded movements are already loaded, they will have short RTs regardless of the go cue delay duration. But, for movements to non-rewarded targets, when the rats are given a short delay between movement and go cues, they do not have enough time to release the DMP and select the new one, so they will be slow. The longer the delay between the cues, the more time the rats have to update to a new plan, allowing them to be quicker to initiate the movement. We saw that the large majority of the rats had a negative correlation between the go cue delay duration and their reaction times (RT) for non-rewarded trials, but no effect for rewarded movement. A consequence of the two contexts and the DMP, is that their interaction could allow for quick adaptation, as an unexpected result implies a change in context. In our setting, with two discrete and deterministic context, observing that one of the targets has changed in value, implies that the other also did. Thus, after a surprising observation, if they are aware of this—as they seem to be—they should change their behavior towards the non observed one. Rats present partial evidence of inference-like behavior in their responses. After a surprising reward, if later sent to the other side, previously rewarded, they already know that they are not going to receive a reward, initiating slowly; but, the in the other case, if they did not receive a reward where they expected to get one, and are later sent to the other side, they go there expecting something different, but not fully committed to a reward. We also observe in the way that animals initiate the trials that they are embodying their knowledge about the context that they are in. Although animals have different strategies, they present different orientation for trials in different contexts, a tattletale of a DMP. Crucially, by our modification in the delay between the cues, we gave the animals a way to express their behavior in a richer manner, this allowed us to see the default motor plan and the effects of their update.

We consider that our behavioral results might drive advantageous steps in our field. To begin, our work presents to the scientific community a new behavioral paradigm, and characterize some of the behavioral measures that can be used to understand the way in which different brain regions or neuromodulatory systems re-

late to them. We show that rats can acquire the underlying structure of the task in the same way that primates can. Whilst, there are noteworthy differences in the way that rats update their responses after a transition with respect to primates. Primates have shown to be capable of inference-like behavior in the 1DR task (Bromberg-Martin et al., 2010b), and update symmetrically to both experienced and non-experienced targets regardless of the value of the outcomes. In our case, we notice that rats are not symmetrical in their updates. For rats, after a surprising rewarding outcome, both subsequent movements are updated; but, after a surprising non-rewarded result, responses are partially updated, in a manner that could be represented as if they were expecting to get something on both sides. The fact that after a surprising non-rewarded result, the distribution of both subsequent RT, towards rewarded or non-rewarded targets, are similar between them, implies that they are being over-optimistic about their environment. We did not foresee this results, as it implies a different underlying process for context updating in rodents, where rewarding outcomes guide quicker updates than non-rewarding ones. One possible explanation is that this can be evolutionary advantageous, if things did not go as good as planned once, it does not imply that it will stay going that way. So, bad outcomes are weighed less heavily in the updating process. This strategy has been used in normative models in reinforcement learning (RL) and has showed to relate to behavioral and neural responses in "bandit" cases (Cazé and van der Meer, 2013). An alternative explanation for the different results with the primate literature, animal model cognitive capacities aside, could be the task structure itself. Generally, in the 1DR tasks the number of blocks is constant, or sampled from few possible values, so animals could use a counting strategy to help themselves in the inference of the change (Kawagoe et al., 1998; Sato and Hikosaka, 2002; Lauwereyns et al., 2002; Bromberg-Martin et al., 2010b); and, in other tasks that have shown evidence of inference-like behavior, animals also receive information about transitions, that facilitate the update of their behavior (Saez et al., 2015). As these task properties, could also make the update of responses more expedite, and can not be disregarded when considering our results.

4.3 Electrophysiological signatures of expectation and context

In chapter 3 we present evidence of movement direction, value, trial context, and interaction between these features or task dimensions in the activity of both SNr and VA/VL. Single cells and population activity analyses of both regions, show that

signals indicative of the block-structure allows for context signals to be present before trial initiation; and, that the integration of this signal with the movement cue related activity, allows the development of the value signal. The population level results of the neural activity across time, indicates that both regions have different levels of stability of the information they convey in dependence on the trial types. These patterns of information stability in the population during the delay before the go cue presentation, relate to the roles that both regions have in the control of behavior within the CBGTC loop.

We found relevant signatures of the task dimensions in both regions at the single cell and population levels. At trial initiation, we found cells informative about the context in both regions, a crucial signature needed to prepare a DMP as this information would allow biasing activity to prepare a movement towards the rewarded location. This early signal was also present in the population analyses. After the movement cue presentation, we encounter in both regions, direction and value related neurons. Finally, at port out, at the two levels of analyses and in both regions, we found direction and value signals. With respect to the evolution of these signals over time, at the single cell level, we observed that SNr neurons had more neurons that carried multiple signals at different moments, multiplexing information. Meaning that the same cell could carry more one signature during the different trial events. If we now focus on the evolution of these signals at the population level, we first notice that the context signal was decodable reliably in SNr before trial initiation, whereas in MTh, it became accurately readable after initiation. After movement cue presentation, we notice that in both areas the movement direction was decodable earlier than value, suggesting the need for an interaction between context and direction that informs the value of the upcoming movement. At the final step in the preparatory process, the port out event, we found that SNr had earlier decodability for direction and value, with respect to MTh, that reliably encoded value. However, MTh had an overall more accurate decodability for movement direction and value around the time of movement initiation, whereas SNr reliably encoded the direction. To evaluate the stability of information over time, we trained linear decoders on activity after movement initiation but before reaching the target, and evaluated those decoders on the activity before the go cue presentation. We found that the stability of SNr activity to decode direction was lower for contralateral movements than for ipsilateral movements. However, stability to decode value was higher for all trial types. In contrast, MTh had unstable activity during the delay period with respect to the activity at port out when decoding direction. For non-rewarded trials, MTh activity for both target directions was more stable. Overall, these signals and their evolution over time are indicative

of relevant processes needed to commit or update a DMP. The early context signal observed indicates that the animals are tracking the rewarded direction, and thus can inform the brain to load a motor plan. The interaction of this contextual signal and the movement direction, can inform the brain about the congruence between the loaded plan and current task demands. The changes in activity at port out can relate to the initiation of the movement. Finally, the changes in decodability across time during the delay between movement- and go-cue can be understood as the correlates of the commitment or updating processes. Where SNr would be helping to withhold a particular movement, and propagating the expected value signal; and MTh, would receive this information to allow cortical and striatal populations to maintain or update the loaded plan.

The electrophysiological results complement directly our behavioral observations, and represent interesting advances in our knowledge about the circuit role in action selection. Evidence about contextual information, and movement direction in both regions could have been expected, as both regions are known to carry these type of signals (Wang et al., 2021; Inagaki et al., 2022). But, our population decoders indicate that contextual signal interacts with movement direction to inform the animal about value, indicating evidence of an integration process occurring with similar profiles in both regions. We propose that the movement direction neurons can inform cortex and motor circuits about the relation between the original plan and the current contingency, to prepare an update of the original plan if needed. The later value signals, could inform about the energy expenditure to be used in the movement, i.e., the vigor to exert (Shadmehr et al., 2019). At single neuron level we observed multiplexing of information, using one channel to convey more than one signal. This can be a consequence of the number of neurons in the BG output being two to three orders of magnitude less than in the striatum, which has been implicated in a dimensionality reduction (Oorschot, 1996; Bar-Gad et al., 2003). But, this could point to a dynamic readout in the receiving neurons. If the same neuron can convey different information, the receiving population needs to be capable of sorting out what is the relevant dimension being informed each time, or use the available information differently. Conveying different information to different population was already observed in MTh in Pavlovian conditioning (Li et al., 2016), but to our knowledge only proposed for SNr (Basso and Sommer, 2011). The presented stability from the population decoders, support that both regions go through fast adaptations, and can relate to this same process of multiplexing information, as they could help to quickly engage regions in more adaptive manners. The difference in time of better than chance decodability of movement and value signals between SNr and MTh at movement cue presentation

are relatively small. But they are in consonance with previous studies showing that neurons in VA nucleus encoding value are also later than SNr (Yasuda and Hikosaka, 2018). Nonetheless, we found larger differences between the regions decodability of movement direction at the initiation of movement, with SNr being earlier than MTh. This goes in line with the role of SNr in motor control via inhibitory projections, but is a novel observation. Finally, our analyses about the stability of the signals over time allowed us to further support the assignment of complementary roles for the region during the fixation period when animals need to either commit or update their plan (Foster et al., 2021; Wang et al., 2021; Inagaki et al., 2022). Finally, we observe that SNr has a prevalent role in the inhibition of the ipsiversive movements, and informing about value throughout the period. Whereas, MTh was mainly stable for non-rewarded movements, a role that is consistent with the thalamocortical and thalamostriatal projections enforcing downstream regions to engage in new patterns of activity, facilitating the update of a plan.

4.4 Proposed synthesis of our results

The behavioral and brain correlates of the responses that rats have in the context of our task are compatible with many implementations. Here, we present two proposals that we consider fruitful for future explorations of our task, and in a broad sense to help understand the way that the BG can communicate to the cortex. We first present a visual depiction of an idealized linear accumulation to threshold model that should behave in manners compatible with our results. We describe how the internal components of this model encompass the relevant signals present in our recordings. After this, we proceed to present a more abstract and general approach, where we show the results of an end-to-end RL model agent and environment that recapitulates some relevant features of the behavior and brain signals. Both proposals are still work in progress and are relevant pieces for our proposals for future work.

4.4.1 A linear accumulation to threshold model

During the task, rats were presented with two different contexts, each with similar behavioral demands but contrasting outcomes for their responses. In each context, only one of the two sides was associated with rewards, making it more appealing to approach. However, due to the penalty for incorrect responses or broken fixations, rats learned to follow the movement cue and perform the requested action to maximize their total rewards, even if this meant going towards the non-rewarded location. The

rats still expressed their intention to go towards the rewarded location in the RTs they used for rewarded and non-rewarded movements; and, in the trials, they choose to abort after being informed about the target location. We showed that this effect can be related to a default motor plan, observable in the side choices that rats made before being informed about the target movement. But also, in the way that the delay until the go cue relates to the RTs for rewarded and non-rewarded trials. Given the observed results, and our discussed interpretation, we consider appropriate to describe the underlying process controlling the behavioral phenomena as an accumulation to threshold model. The LATER model family (Noorani and Carpenter, 2016) proposes that in tasks not dependent on evidence collection or integration (Carpenter and Reddi, 2001), a linear accumulation to threshold process can model the observed effect in movement initiation once different cues are presented. The model has been extended to account for trial to trial variability in behavioral responses that can be related to both pre- or post-execution processes within an experimental block or trial(Nakahara et al., 2006). This family of models includes the effects of the accumulation starting point on RT distribution, and the variance of the accumulation process could relate to vigor, which are plausible tools to be used to express our results. Taking this general framework into account, we can summarize our behavioral observation by means of the diagram presented in panel A from figure 4.1. In the diagram, we express the behavior of an agent as a linear accumulation to threshold process with two boundaries. If the process reaches the top, a movement to the right is initiated, if in the bottom, one to the left. In the example, the agent is in a block where reward would be given for rightward movements. Consequently, the initial state of this accumulation is biased towards this location, setting a context bias. Before the go cue delivery, an inhibitory process must suppress the decision variable from reaching the boundaries, visualized as the red/yellow underlines during the whole period leading to the go cue presentations, when the process is allowed to diffuse. In case of failure to inhibit the diffusion and the process were to reach a bound, there would be a broken fixation, more likely leading to a movement to the biased direction. After a movement cue is presented, it must be recognized through a perceptual process. If the cue informs that the movement direction is to the biased side, the initiation inhibition must be maintained. But, if the cue informs that the target direction is to the non-preferred location, the system needs to move away from the contextual bias, to accommodate for the new target. To this end, the process needs to be disinhibited in a controlled manner, weakening the suppression of diffusion initiation, which could lead to higher broken fixation rates in this condition. After go cue delivery, for movements toward the rewarded location, the decision variable at similar distances of the boundary regardless of delay duration,

so the diffusion process must take comparable times to reach the bound. In the complementary case, non-rewarded movement, the process needs to update, hence for short go cue delays, the diffusion takes longer to reach the boundary, and the longer the delay, the quicker it can be reached. Given the differences associated to outcome value, we include different angles of the accumulation once the process is allowed to drift. For rewarded movements the angle is steeper, α, allowing reaching quickly the boundary; non-rewarded movements have a shallower angle, β, which leads to slower reaction times even in case that the process were to reach the level of context-bias associated to the other context, not shown in the sketch.

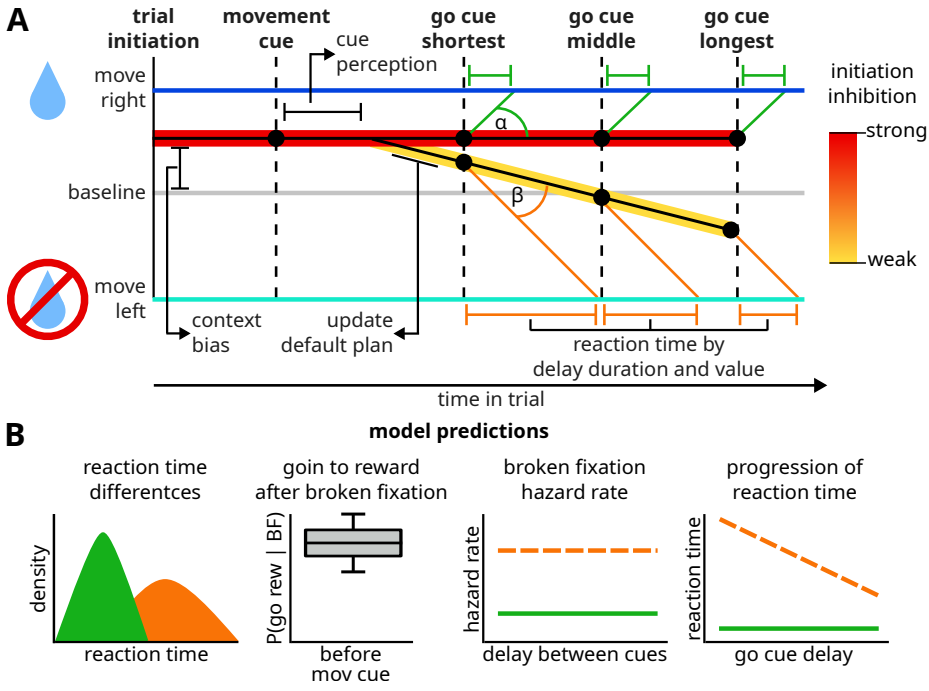


Figure 4.1: Graphical depiction of the proposed a behavioral controller and predictions. **A.** Depiction of an accumulation to threshold model to explain our behavioral observations. A decision variable accumulates until it reaches one of two boundaries, after this it initiates the action associated to that bound. The explanation of the diagram is given in the text. **B.** Predictions derived from the model: Given a contextual bias, the model predicts effects in reaction times by target value (leftmost panel). The bias also must affect choices for broken fixation before movement cue (center-left panel). In case of plan update, the model predicts different rates of broken fixation by target value (center-right panel). The duration of the go cue delay must influence the progression of reaction times by target value (rightmost panel). In the sketches we use *blue* for rightward movements, and *light-blue* for left; *green* lines are rewarded movements, and *orange* non-rewarded movements; finally, *red* and *yellow* represent strong or weak inhibition of movement initiation.

From this model, we can derive predictions related to RT differences depending on the movement value, effects over broken fixation before and after movement cue presentation, and effects over RT related to changes of plans (Fig. 4.1, panel B).

The first prediction (leftmost panel) is that the context bias implies a difference in the distribution of RTs for rewarded and non-rewarded movements, as the process should reach the boundary associated to reward quicker than for non rewarded. The controlled suppression of the inhibition of diffusion initiation for non-rewarded target movements implies that these trials should have a higher rate of broken fixations throughout the delay between the movement and go cue (center-left panel). The model predicts also that for broken fixations before the movement cue, the process should be more likely to reach the rewarded boundary, as the context bias has the process closer to this bound (center-right panel). Finally, the controlled suppression of diffusion inhibition entails that for rewarded movements, the process should reach the boundary at similar times; whereas for non-rewarded movements, short go cue delays require a longer diffusion period, but the longer the delay, the distance between the initiation and the boundary is shortened (leftmost panel). A consequence of α & β, is that differences in RT by value will stay present in the limit of the process initiating from the same context bias, as the angles will drive short RTs for rewarded and longer for non-rewarded.

The prediction derived from the model are compatible with the observed behavioral results. And, we can expect that the underlying dynamics of the model components could relate to our electrophysiological data. The model requires the maintenance of a context bias signal throughout the block, whilst also needs to allow updating the inhibitory process if the task demands a non preferred movement. Our results indicate that SNr maintains contextual information reliably throughout the trial, whereas VA/VL does this transiently at trial initiation and early after movement cue presentation. These observations support the notion that the contextual bias signal can be enforced by SNr and MTh early in the trial, and the persistence in SNr decodablity could track the context over the block duration. The model includes a period between movement cue presentation and the use of the information to update the decision variable location. The linear readouts from our recordings show that both regions include signatures of this information, where there is a delay between movement cue presentation, and the increase in decodability for direction and value. From the results of our decoders across time, off-diagonal decoding results, we can relate SNr activity to the facilitation of the suppression of action initiation. Whilst VA/VL shows patterns that can advance the update of a plan given non-preferred demands. The quick rise in decodability across time of SNr in the value dimension for non-rewarded movements, can relate to the different angles that the model would use to drift towards the targets depending on expected value. For all conditions, quickly after cue presentation and before go cue is presented, SNr is capable of organizing his

activity to inform reliably other regions about the value of the upcoming movement. This would allow non-rewarding movements to be taken with less vigor, and rewarding ones more readily. Similarly, both regions quickly rise in decodability for direction and value before port out, with SNr being earlier than MTh in direction. This can indicate that the former region is quickly informing other areas about the upcoming action to perform, in particular the latter area, whilst both could inform about the expected outcome.

In summary, our model is capable of describing and predicting parts of the behavioral phenotype that rats display, and the dynamics observed in our recordings are compatible with the internal changes required for it to operate. Even though not implemented yet, pursuing the deployment of this model could be informative for the field. As it can be used to make new predictions about behavioral responses to task manipulations, and relate brain activity to the underlying deliberation processes.

4.4.2 An end-to-end RL agent and environment

In the framework of RL, agents interact with environments by taking actions, depending on the state of the environment those actions will modify the environment state. And at each interaction the environment gives an observation and some reward value. The agents can learn to interact with the environment in different way, by learning a mapping of states and actions, policy function, or an action-value pairings, value functions. The actor-critic algorithms are model-free agents, where the **actor** learns the optimal action to take, policy function, whilst the **critic** evaluates how good is the performance of the current policy, value function (Sutton and Barto, 2018). Both functions are optimized in parallel by interacting with an environment, comparing the obtained results with an expected outcome via a temporal discount (TD) error (Fig. 4.2 A). New implementations of actor-critic algorithms have used deep neural networks to instantiate function approximators that learn to respond adaptively in untrained environments (Botvinick et al., 2020). With the help, support, and patience of a colleague from the lab, we implemented a version of the asynchronous-advantage-actor-critic (A3C) architecture proposed by Mnih et al. (2016). This model has been interpreted as an abstraction of the cortico-basal ganglia-thalamo-cortical (CBGTC) loop (Wang et al., 2018) (Fig. 4.2 B top left). The state representation learned by the model is paired with the prefrontal cortex, while the TD error is related to DA activity. The BG output and MTh represent the action and value estimates, and the perception of current observation, previous actions, and outcomes are all part of the cortex and BG. As the model shows interesting features in terms of state represen-

tations and generalization capabilities. The model internally uses a long-short-term memory (LSTM) cell, a recurrent neural network architecture capable of learning long range relation between elements in the input (Fig. 4.2 B top center panel). The LSTM learns what are elements are relevant to maintain from the received input, and which to forget. These elements allow the model to learn a general state representation that is capable of responding to new environments (Wang et al., 2018; Botvinick et al., 2020). The LSTM input output are graphically depicted as a box diagram (Fig. 4.2, B top right). The network receives as input from the environment the current observation, and from itself the previous action and outcome. The input is evaluated by the RNN and returns a policy by a softmax over the readout, a distribution of action probabilities, and a value estimate, from a linear readout. From the policy distribution, one action is selected by sampling the probabilities. Importantly, in Wang et al. (2018), the researchers share the same weights for both the policy and value functions, in contrast to Mnih et al. (2016). The objective used to learn takes into account the sum of the policy gradient, the state-value function loss (using advantage instead of direct returns), and an entropy regularization term to allow for exploration in action space (Fig. 4.2, B bottom).

We first developed an environment that behaved in a manner similar to our delay movement task (Fig. 4.2, C). The environment is a discrete-time finite-state-machine that emits as observation an array of three integer values, representing the three nose-ports, and receives one of three possible actions. The observation array elements and actions are interpretable as left, right, and center port; or the action of going to the left, center, or right port respectively. After each interaction the time advanced one time unit and depending on the action taken, the state machine updated his internal state, emits the values of the updated observation array and an outcome value of the previous interaction. At the beginning of an epoch, the state machine selects one of the two edges of the observation array as the reward location, and a number of trials for the current block. In each trial, the state machine starts at state *initiation*, here a movement direction is randomly selected from the two edges of the observation array. The *initiation* state emits the observation array with a cue value at the center position. If the agent uses the center action, the machine transitions into the *movement cue* state, else it stays in the same. In the *movement cue* state, the observation array displays a cue value on the chosen movement direction, taking the center action would advance the state machine to the *go cue* state. In this state, the observation array also displays the cue value at the movement direction. If the agent takes the center action in the *go cue* state, it would continue in the same state, and taking the action associated to the observation cue would transition to the state *feedback*. If the agent

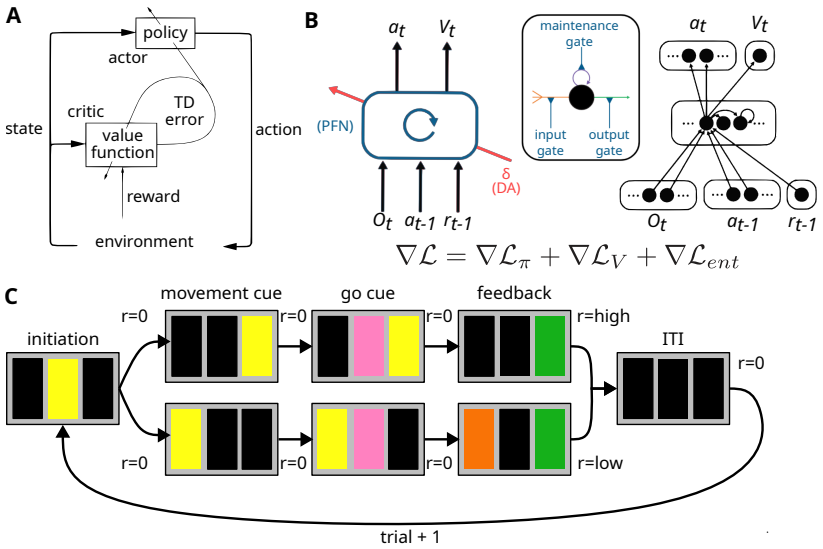


Figure 4.2: An actor-critic in a simplified delayed movement task. **A.** Graphical depiction of the actor-critic algorithm, an environment presents a state and a reward to an agent that takes actions using the policy, actor, that is evaluated by their outcomes with respect to what was expected, critic. The agent learns by minimizing the TD error a policy that leads to the higher value estimates. Adapted from Sutton and Barto (1998). **B.** Visualization of an implementation of the actor-critic algorithm using a recurrent neural network. The left panel shows how the model relates to the CBGTC loop. The center panel depicts the LSTM cell at the core of the network. The right panel makes more explicit the inner workings of the LSTM architecture. The bottom panel describes the loss function. Abbreviations include a for action, V for value, O for observation and r for reward, the subscripts indicate current or previous time Adapted from Wang et al. (2018). **C.** Our delayed movement task as an environment to be used by an agent, the example presents a block where rewards would be given for rightward movements.

does not take the center action during the *movement cue*, or takes the incorrect action at the *go cue* state, the environment moves to an *error* state. The *error* state emits an empty observation array for 5 time steps, the duration of a trial in discrete time, regardless of actions taken. If the state *feedback* has been reached, and the movement cue and reward location for the current trial were congruent, the environment gives a large reward. In case they are incongruent, the environment gives a small reward. Any action taken in the *feedback* state advances the environment to an *inter-trial-interval* state that advances to a trial counter one value, and a new trial initiates, where the cycle would continue. All non-*feedback* states give no rewards. After the number of trials for the block had passed, the reward location flips to the other location, and a new number of trials for the block is sampled. The epoch ends after 200 trials, or the equivalent number of time steps. Our agent was implemented as an asynchronous-advantage actor-critic model (Mnih et al., 2016; Wang et al., 2018), for the internals of the agent, we use a LSTM cell with 64 recurrently connected units. In our implementation, as in Wang et al. (2018), we use the same internal weights for the policy and value estimators, which are implemented as a softmax and linear readouts of the LSTM internal state respectively. As the vigor literature has shown, there is an inverse relation between expected value and RT (Shadmehr et al., 2016, 2019), taking this into account, we use the inverse of the value times some constant to translate the agent value estimates into RTs.

In our implementation, a vanilla A3C, the agent learns sensitive value and policy functions (Fig. 4.3, A). With respect to the policy, we observe that the agent learns to select the appropriate action depending on trial condition and environment state (Fig. 4.3, A, top 3 panels). The agent selects consistently the center action for all trial conditions in all non *go cue* states, and depending on trial laterality, the agent selects the correct action. We observe that the agent accurately predicts the outcomes for each environment state, and the expected value starts increasing for highly rewarded trials before the reward is given (Fig. 4.3, A bottom panel). The agent displays distribution of RT compatible with the animals, quick for rewarded and slow for non rewarded (Fig. 4.3, B). But unlike the rats, the agent achieves perfect inference, after transitions the agent updates his behavior after one experience, and does so symmetrically regardless of target value (Fig. 4.3, C). The agent rarely makes broken fixations or errors (not depicted), thus we can not compare these results. Even though there are discrepancies observed between the model and the animals, the general framework looks like a promising approach to further understand the underlying processes related to adaptive behavioral control. Many of the underlying

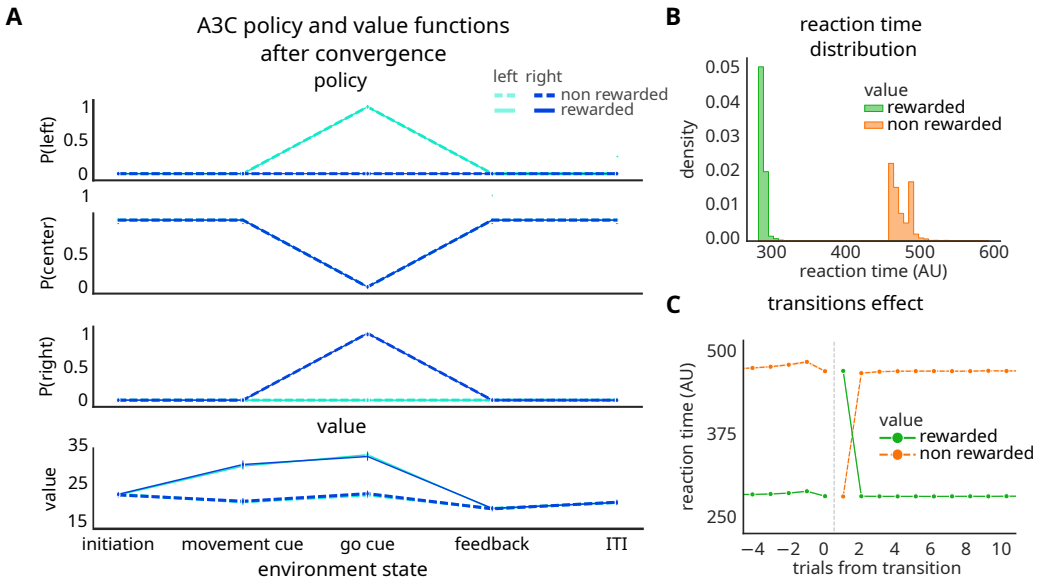


Figure 4.3: Vanilla A3C agent learns a sensitive policy and value function, and behaves optimally. A. Policy and value functions learned by the A3C agent after being trained. All panels share in the x-axis the different states. The P(*side*) axes indicate the probabilities associated to each possible action at each state. The value axis indicates the expected value assigned to each state. The coloring and line styles follow the schema in figure 3.1, panel C. **B.** The agent uses different distribution of RTs for trials depending on the expected value. **C.** The agent updates symmetrically for both types of trials after a block transition. Panels B & C use *green* for rewarded, and *orange* for non-rewarded.

properties of the agent and the environment can be manipulated to evaluate and propose new hypothesis about the ways in which the CBGTC loop works.

4.5 Limitations of our work and future plans

Experimental work is critical for advancing scientific knowledge. But, as any other human endeavor, our work presents limitations that tamper the possible analyses we were able to perform on the data, and confine the scope and the interpretation of our results. Here we are going to focus on some of the limitations that we observed, and suggest ways to mitigate them in future work.

One problem that we did not realize until late in the data acquisition process was that our time limits for response, allowed animals to stay idle for too long and still make a correct response. It may have been possible to gather even more trials and more transitions if we were to use a shorter maximum response delay. We also used too large block sizes, and sampled the range available uniformly. Even though this did not lead to observable prediction of block transitions in rats responses, it also

lowered the number of transitions that we got per session. Future iterations of the experimental paradigm should take these issues into account, to maximize the amount of relevant epochs needed to characterize the contextual update within the task.

Aside from the difficulties in the implementation or parameters used in the task itself, our results are also incapable of resolving key issues needed to relate the observed activity of the recorded regions and the BG to cortex communication. None of our observations are capable of responding if these signals are inherited from BG or cortical activity, or which signals are sent to any of the downstream regions. Hence, the finer-grain details of the particular roles of BG computations over cortical activity are still amiss. Future experiments could record simultaneously more regions with newly available and developing technologies that allow recordings of hundreds to thousands of neurons across different brain regions and cerebral axes (Jun et al., 2017; Juavinett et al., 2018; van Daal et al., 2021). Another possibility is to use optogenetic tools in tandem with our recording setup to label cells in SNr or MTH that are receiving-from or projecting-to particular striatal or cortical territories. This would allow to directly observe the particular information being multiplexed between the areas (Lima et al., 2009). Moreover, these optogenetic tools could allow manipulating the activity of our neuronal population, giving us access to manipulations that could inform about the causal roles of the observed signatures in behavior.

With respect to the models proposed as synthesis of our work, both are still in the earlier stages of development, and this is why we only present them as appetizing proposals and not fully developed chapters by themselves. On the one hand, the linear accumulator model is still missing an appropriate implementation, the development of this would highly enrich future experimental designs and hypothesis. On the other, our RL model, even though in better shape, is still missing relevant behavioral and internal dynamics. These effects can be related to the environment not being rich enough to allow the agent to express these properties, thus by allowing for different delays durations between the trial events, we could make the environment closer to our task. Another possibility is that our agent itself could be missing relevant properties. To mitigate this issue, and taking a clear position about the relevant features that the agent would need. We could take into consideration the *many worlds hypothesis* for DA (Lau et al., 2017), that proposes that the parallel loops in the CBGTC circuit give rise to parallel agents that have access to sparse levels of information. The final action taken and value expected from the organism at large are the result of a pooled voting between these multiple independent agents. To implement this idea, we could sample few units from the LSTM and use their representations to guide different policy and value functions. And use a pooled version of these proposals as the agent policy and

value estimates. This formulation presents as a very fruitful endeavor, as end-to-end models allows putting to test broader hypothesis about the mechanisms underlying behavioral control.

4.6 Conclusion

From our foray into the behavior and brain correlates of the rats responding to our delayed movement task, we already described many interpretations of the results. We took a very simple behavioral task, that has very little space for animals to express in their creative ways how to solve the problems that we propose. Yet, even in this confined world, we were surprised to find how much we could learn about them by careful observation of their responses. Besides our behavioral results, the most clear take-away from our exploration of the neural correlates of our task is that neither SNr nor MTh have simple *context*, *movement*, or *value* neurons. Both regions show a very rich and dynamical myriad of patterns that are propagated to their downstream receiving areas, many of which will quickly give back their own activity into the CBGTC loop. Either in the form of a movement in execution, a change in expectation, or some perturbation in the high dimensional space of cortical activity. The fact that we found this rich dynamics in areas so conserved across species should be read as a relevant feature of the internal properties of the system. Brains use their dynamics to encode and manipulate information, we are still in the early days of exploring the rules that govern the trajectories in the available space of possible configurations. And, for the future explorers, we give a simple vessel that can help to navigate movement preparation and anticipation in context driven behavioral control.

References

Aarts, H. and Elliot, A. J., editors (2012). *Goal-Directed Behavior*. Frontiers of Social Psychology. Psychology Press, New York, NY.

Administration (US), S. A. a. M. H. S. and General (US), O. o. t. S. (2016). *THE NEUROBIOLOGY OF SUBSTANCE USE, MISUSE, AND ADDICTION*. US Department of Health and Human Services.

Aflalo, T. N. and Graziano, M. S. A. (2006). Possible Origins of the Complex Topographic Organization of Motor Cortex: Reduction of a Multidimensional Space onto a Two-Dimensional Array. *Journal of Neuroscience*, 26(23):6288–6297.

Ajuwon, V., Ojeda, A., Murphy, R. A., Monteiro, T., and Kacelnik, A. (2022). Paradoxical choice and the reinforcing value of information. *Animal Cognition*.

Akre, K. L. and Johnsen, S. (2014). Psychophysics and the evolution of behavior. *Trends in Ecology & Evolution*, 29(5):291–300.

Albaugh, D. L., Huang, C., Ye, S., Paré, J.-F., and Smith, Y. (2021). Glutamatergic inputs to GABAergic interneurons in the motor thalamus of control and parkinsonian monkeys. *The European Journal of Neuroscience*, 53(7):2049–2060.

Alexander, G. E., DeLong, M. R., and Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, 9:357–381.

Alvarez, E. O. and a Alvarez, P. (2008). Motivated exploratory behaviour in the rat: The role of hippocampus and the histaminergic neurotransmission. *Behavioural brain research*, 186(1):118–25.

Antal, M., Beneduce, B. M., and Regehr, W. G. (2014). The Substantia Nigra Conveys Target-Dependent Excitatory and Inhibitory Outputs from the Basal Ganglia to the Thalamus. *Journal of Neuroscience*, 34(23):8032–8042.

Aoki, S., Smith, J. B., Li, H., Yan, X., Igarashi, M., Coulon, P., Wickens, J. R., Ruigrok, T. J., and Jin, X. (2018). An open cortico-basal ganglia loop allows limbic

control over motor output via the nigrothalamic pathway. *eLife*, 8.

Armstrong, K. M., Chang, M. H., and Moore, T. (2009). Selection and Maintenance of Spatial Information by Frontal Eye Field Neurons. *Journal of Neuroscience*, 29(50):15621–15629.

Balleine, B. W. and O'Doherty, J. P. (2010). Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. *Neuropsychopharmacology*, 35(1):48–69.

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

Basso, M. A., Pokorny, J. J., and Liu, P. (2005). Activity of substantia nigra pars reticulata neurons during smooth pursuit eye movements in monkeys. *The European Journal of Neuroscience*, 22(2):448–464.

Basso, M. A. and Sommer, M. A. (2011). EXPLORING THE ROLE OF THE SUBSTANTIA NIGRA PARS RETICULATA IN EYE MOVEMENTS. *Neuroscience*, 198:205–212.

Basso, M. A. and Wurtz, R. H. (2002). Neuronal Activity in Substantia Nigra Pars Reticulata during Target Selection. *Journal of Neuroscience*, 22(5):1883–1894.

Benhamou, L. and Cohen, D. (2014). Electrophysiological characterization of entopeduncular nucleus neurons in anesthetized and freely moving rats. *Frontiers in Systems Neuroscience*, 8.

Bennett, M. R. and Hacker, P. M. S. (2012). Perceptions, Sensations and Cortical Function: Helmholtz to Singer. In *History of Cognitive Neuroscience*, pages 4–43. John Wiley & Sons, Ltd.

Bentivoglio, M., Spreafico, R., Minciacchi, D., and Macchi, G. (1991). GABAergic interneurons and neuropil of the intralaminar thalamus: An immunohistochemical study in the rat and the cat, with notes in the monkey. *Experimental Brain Research*, 87(1):85–95.

Berlyne, D. E. (1955). The arousal and satiation of perceptual curiosity in the rat. *Journal of Comparative and Physiological Psychology*, 48:238–246.

Berridge, K. C. (2007). The debate over dopamine's role in reward: The case for incentive salience. *Psychopharmacology*, 191(3):391–431.

Bertelson, P. (1967). The Time Course of Preparation*. *Quarterly Journal of Experimental Psychology*, 19(3):272–279.

Bickford, M. E. (2016). Thalamic Circuit Diversity: Modulation of the Driver/Modulator Framework. *Frontiers in Neural Circuits*, 9:86.

Bosch-Bouju, C., Hyland, B. I., and Parr-Brownlie, L. C. (2013). Motor thalamus integration of cortical, cerebellar and basal ganglia information: Implications for normal and parkinsonian conditions. *Frontiers in Computational Neuroscience*, 7:163.

Botvinick, M., Wang, J. X., Dabney, W., Miller, K. J., and Kurth-Nelson, Z. (2020). Deep Reinforcement Learning and Its Neuroscientific Implications. *Neuron*, 107(4):603–616.

Bouton, M. E. (2021). Context, attention, and the switch between habit and goal-direction in behavior. *Learning & Behavior*, 49(4):349–362.

Boyd-Meredith, J. T., Piet, A. T., Dennis, E. J., El Hady, A., and Brody, C. D. (2022). Stable choice coding in rat frontal orienting fields across model-predicted changes of mind. *Nature Communications*, 13(1):3235.

Brandeis, R., Brandys, Y., and Yehuda, S. (1989). The use of the Morris Water Maze in the study of memory and learning. *The International Journal of Neuroscience*, 48(1-2):29–69.

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

Bromberg-Martin, E. S., Matsumoto, M., Hong, S., and Hikosaka, O. (2010b). A Pallidus-Habenula-Dopamine Pathway Signals Inferred Stimulus Values. *Journal of Neurophysiology*, 104(2):1068–1076.

Bruce, C. J. and Goldberg, M. E. (1985). Primate frontal eye fields. I. Single neurons discharging before saccades. *Journal of Neurophysiology*, 53(3):603–635.

Bryden, D. W., Johnson, E. E., Diao, X., and Roesch, M. R. (2011). Impact of expected value on neural activity in rat substantia nigra pars reticulata. *European Journal of Neuroscience*, 33(12):2308–2317.

Bundt, C., Abrahamse, E. L., Braem, S., Brass, M., and Notebaert, W. (2016). Reward anticipation modulates primary motor cortex excitability during task preparation. *NeuroImage*, 142:483–488.

Bundt, C., Bardi, L., Verbruggen, F., Boehler, C. N., Brass, M., and Notebaert, W. (2019). Reward anticipation changes corticospinal excitability during task preparation depending on response requirements and time pressure. *Cortex*, 120:159–168.

Butz, M. V., Herbort, O., and Pezzulo, G. (2008). Anticipatory, Goal-Directed Behavior. In Pezzulo, G., Butz, M. V., Castelfranchi, C., and Falcone, R., editors, *The Challenge of Anticipation: A Unifying Framework for the Analysis and Design of Artificial Cognitive Systems*, Lecture Notes in Computer Science, pages 85–113. Springer, Berlin, Heidelberg.

Calaminus, C. and Hauber, W. (2009). Modulation of behavior by expected reward magnitude depends on dopamine in the dorsomedial striatum. *Neurotoxicity Research*, 15(2):97–110.

Callaway, J. K., Jones, N. C., and Royse, C. F. (2012). Isoflurane induces cognitive deficits in the Morris water maze task in rats. *European Journal of Anaesthesiology*, 29(5):239–245.

Capizzi, M., Visalli, A., Faralli, A., and Mioni, G. (2022). Explicit and implicit timing in older adults: Dissociable associations with age and cognitive decline. *PloS One*, 17(3):e0264999.

Carpenter, R. H. S. and Reddi, B. a. J. (2001). Reply to 'Putting noise into neurophysiological models of simple decision making'. *Nature Neuroscience*, 4(4):337–337.

Catanese, J. and Jaeger, D. (2021). Premotor Ramping of Thalamic Neuronal Activity Is Modulated by Nigral Inputs and Contributes to Control the Timing of Action Release. *Journal of Neuroscience*, 41(9):1878–1891.

Çavdar, S., Özgür, M., Uysal, S. P., and Amuk, Ö. C. (2014). Motor afferents from the cerebellum, zona incerta and substantia nigra to the mediodorsal thalamic nucleus in the rat. *Journal of Integrative Neuroscience*, 13(04):565–578.

Cazé, R. D. and van der Meer, M. A. A. (2013). Adaptive properties of differential learning rates for positive and negative outcomes. *Biological Cybernetics*, 107(6):711–719.

Centonze, D., Picconi, B., Gubellini, P., Bernardi, G., and Calabresi, P. (2001). Dopaminergic control of synaptic plasticity in the dorsal striatum. *The European Journal of Neuroscience*, 13(6):1071–1077.

Chance, P. (1999). Thorndike's Puzzle Boxes And The Origins Of The Experimental Analysis Of Behavior. *Journal of the Experimental Analysis of Behavior*, 72(3):433–440.

Chelazzi, L., Rossi, F., Tempia, F., Ghirardi, M., and Strata, P. (1989). Saccadic Eye Movements and Gaze Holding in the Head-Restrained Pigmented Rat. *The European Journal of Neuroscience*, 1(6):639–646.

Chen, Z., Zhang, Z.-Y., Zhang, W., Xie, T., Li, Y., Xu, X.-H., and Yao, H. (2021). Direct and indirect pathway neurons in ventrolateral striatum differentially regulate licking movement and nigral responses. *Cell Reports*, 37(3):109847.

Chenouard, V., Remy, S., Tesson, L., Ménoret, S., Ouisse, L.-H., Cherifi, Y., and Anegon, I. (2021). Advances in Genome Editing and Application to the Generation of Genetically Modified Rat Models. *Frontiers in Genetics*, 12.

Choi, J. E. S., Vaswani, P. A., and Shadmehr, R. (2014). Vigor of Movements and the Cost of Time in Decision Making. *Journal of Neuroscience*, 34(4):1212–1223.

Cisek, P. (2007). Cortical mechanisms of action selection: The affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1485):1585–1599.

Crego, A. C. G., Štoček, F., Marchuk, A. G., Carmichael, J. E., van der Meer, M. A. A., and Smith, K. S. (2020). Complementary Control over Habits and Behavioral Vigor by Phasic Activity in the Dorsolateral Striatum. *The Journal of Neuroscience*, 40(10):2139–2153.

Crittenden, J. R. and Graybiel, A. M. (2011). Basal Ganglia disorders associated with imbalances in the striatal striosome and matrix compartments. *Frontiers in Neuroanatomy*, 5:59.

Cruz, B. F., Guiomar, G., Soares, S., Motiwala, A., Machens, C. K., and Paton, J. J. (2022). Action suppression reveals opponent parallel control via striatal circuits. *Nature*, 607(7919):521–526.

de Bruin, J. P., Swinkels, W. A., and de Brabander, J. M. (1997). Response learning of rats in a Morris water maze: Involvement of the medical prefrontal cortex. *Behavioural Brain Research*, 85(1):47–55.

Dehaene, S. and Changeux, J. P. (2000). Reward-dependent learning in neuronal networks for planning and decision making. *Progress in Brain Research*, 126:217–229.

Didden, R., Sigafoos, J., O'Reilly, M. F., Lancioni, G. E., and Sturmey, P. (2007). A Multisite Cross-Cultural Replication of Unsuccessful Self-Treatment of Writer's Block. *Journal of Applied Behavior Analysis*, 40(4):773.

Ding, L. and Hikosaka, O. (2007). Temporal Development of Asymmetric Reward-Induced Bias in Macaques. *Journal of Neurophysiology*, 97(1):57–61.

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, 30:412–431.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Number 57 in Monographs on Statistics and Applied Probability. Chapman & Hall, New York.

Ennaceur, A. and Delacour, J. (1988). A new one-trial test for neurobiological studies of memory in rats. 1: Behavioral data. *Behavioural Brain Research*, 31(1):47–59.

Erlich, J. C., Bialek, M., and Brody, C. D. (2011). A cortical substrate for memory-guided orienting in the rat. *Neuron*, 72(2):330–343.

Everitt, B. J. and Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, 8(11):1481–1489.

Fields, H. L., Hjelmstad, G. O., Margolis, E. B., and Nicola, S. M. (2007). Ventral tegmental area neurons in learned appetitive behavior and positive reinforcement.

Annual review of neuroscience, 30:289–316.

Florio, T. M., Scarnati, E., Rosa, I., Di Censo, D., Ranieri, B., Cimini, A., Galante, A., and Alecci, M. (2018). The Basal Ganglia: More than just a switching device. *CNS neuroscience & therapeutics*, 24(8):677–684.

Foddy, B. (2016). Addiction: The pleasures and perils of operant behavior. In Heather, N. and Segal, G., editors, *Addiction and Choice: Rethinking the Relationship*, page 0. Oxford University Press.

Foerde, K. and Shohamy, D. (2011). The role of the basal ganglia in learning and memory: Insight from Parkinson's disease. *Neurobiology of learning and memory*, 96(4):624–636.

Foster, N. N., Barry, J., Korobkova, L., Garcia, L., Gao, L., Becerra, M., Sherafat, Y., Peng, B., Li, X., Choi, J.-H., Gou, L., Zingg, B., Azam, S., Lo, D., Khanjani, N., Zhang, B., Stanis, J., Bowman, I., Cotter, K., Cao, C., Yamashita, S., Tugangui, A., Li, A., Jiang, T., Jia, X., Feng, Z., Aquino, S., Mun, H.-S., Zhu, M., Santarelli, A., Benavidez, N. L., Song, M., Dan, G., Fayzullina, M., Ustrell, S., Boesen, T., Johnson, D. L., Xu, H., Bienkowski, M. S., Yang, X. W., Gong, H., Levine, M. S., Wickersham, I., Luo, Q., Hahn, J. D., Lim, B. K., Zhang, L. I., Cepeda, C., Hintiryan, H., and Dong, H.-W. (2021). The mouse cortico–basal ganglia–thalamic network. *Nature*, 598(7879):188–194.

Frese, M. and Sabini, J. (2021). *Goal Directed Behavior: The Concept of Action in Psychology*. Routledge, London, first edition.

Friedman, N. P. and Robbins, T. W. (2022). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology*, 47(1):72–89.

Froeberg, S. (1907). *The Relation between the Magnitude of Stimulus and the Time of Reaction*. New York.

Galton, F. (1890). Exhibition of Instruments (1) for Testing Perception of Differences of Tint, and (2) for Determining Reaction-Time. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 19:27–29.

Garcia-Munoz, M. and Arbuthnott, G. W. (2015). Basal ganglia—thalamus and the "crowning enigma". *Frontiers in Neural Circuits*, 9.

Gerfen, C. R. and Bolam, J. P. (2010). Chapter 1 - The Neuroanatomical Organization of the Basal Ganglia. In Steiner, H. and Tseng, K. Y., editors, *Handbook of Behavioral Neuroscience*, volume 20 of *Handbook of Basal Ganglia Structure and Function*, pages 3–28. Elsevier.

Gerfen, C. R., Paletzki, R., and Heintz, N. (2013). 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. and Surmeier, D. J. (2011). Modulation of striatal projection systems by dopamine. *Annual review of neuroscience*, 34:441–466.

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 108 Suppl:15647–54.

Graybiel, A. M. (1995). Building action repertoires: Memory and learning functions of the basal ganglia. *Current Opinion in Neurobiology*, 5(6):733–741.

Graybiel, A. M., Aosaki, T., Flaherty, A. W., and Kimura, M. (1994). The basal ganglia and adaptive motor control. *Science (New York, N.Y.)*, 265(5180):1826–1831.

Graziano, M. S. A. (2009). *The Intelligent Movement Machine: An Ethological Perspective on the Primate Motor System*. Oxford University Press, Oxford ; New York.

Grillner, S. and Robertson, B. (2016). The Basal Ganglia Over 500 Million Years. *Current biology: CB*, 26(20):R1088–R1100.

Gulcebi, M. I., Ketenci, S., Linke, R., Hacıoğlu, H., Yanalı, H., Veliskova, J., Moshé, S. L., Onat, F., and Çavdar, S. (2012). Topographical connections of the substantia nigra pars reticulata to higher-order thalamic nuclei in the rat. *Brain Research Bulletin*, 87(2–3):312–318.

Haber, S. N. and Calzavara, R. (2009). The cortico-basal ganglia integrative network: The role of the thalamus. *Brain Research Bulletin*, 78(2–3):69–74.

Hackley, S. A. (2009). The speeding of voluntary reaction by a warning signal. *Psychophysiology*, 46(2):225–233.

Hamadjida, A., Dea, M., Deffeyes, J., Quessy, S., and Dancause, N. (2016). Parallel Cortical Networks Formed by Modular Organization of Primary Motor Cortex Outputs. *Current Biology*, 26(13):1737–1743.

Handel, A. and Glimcher, P. W. (2000). Contextual Modulation of Substantia Nigra Pars Reticulata Neurons. *Journal of Neurophysiology*, 83(5):3042–3048.

Hardwick, R. M., Forrence, A. D., Costello, M. G., Zackowski, K., and Haith, A. M. (2022). Age-related increases in reaction time result from slower preparation, not delayed initiation. *Journal of Neurophysiology*, 128(3):582–592.

Hart, J. (2015). Higher-Order Sensory Processing. In Hart, John, J., editor, *The Neurobiology of Cognition and Behavior*, page 0. Oxford University Press.

Hauser, C. K., Zhu, D., Stanford, T. R., and Salinas, E. (2018). Motor selection dynamics in FEF explain the reaction time variance of saccades to single targets. *eLife*, 7.

Henderson, A. R. (2005). The bootstrap: A technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. *Clinica Chimica Acta*, 359(1):1–26.

Henderson, L. and Dittrich, W. H. (1998). Preparing to react in the absence of uncertainty: I. New perspectives on simple reaction time. *British Journal of Psychology (London, England: 1953)*, 89 (Pt 4):531–554.

Henry, F. M. and Rogers, D. E. (1960). Increased response latency for complicated movements and a "memory drum" theory of neuromotor reaction. *Research Quarterly of the American Association for Health, Physical Education, & Recreation*, 31:448–458.

Hikosaka, O., Kim, H. F., Yasuda, M., and Yamamoto, S. (2014). Basal ganglia circuits for reward value-guided behavior. *Annual review of neuroscience*, 37:289–306.

Hikosaka, O., Nakamura, K., and Nakahara, H. (2006). Basal Ganglia Orient Eyes to Reward. *Journal of Neurophysiology*, 95(2):567–584.

Hikosaka, O., Takikawa, Y., and Kawagoe, R. (2000). Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiological reviews*, 80(3):953–978.

Hikosaka, O. and Wurtz, R. H. (1983). Visual and oculomotor functions of monkey substantia nigra pars reticulata. III. Memory-contingent visual and saccade responses. *Journal of Neurophysiology*, 49(5):1268–1284.

Hikosaka, O. and Wurtz, R. H. (1985). Modification of saccadic eye movements by GABA-related substances. II. Effects of muscimol in monkey substantia nigra pars reticulata. *Journal of Neurophysiology*, 53(1):292–308.

Hikosaka, O., Yasuda, M., Nakamura, K., Isoda, M., Kim, H. F., Terao, Y., Amita, H., and Maeda, K. (2019). Multiple neuronal circuits for variable object–action choices based on short- and long-term memories. *Proceedings of the National Academy of Sciences*, 116(52):26313–26320.

Hill, D. N., Mehta, S. B., and Kleinfeld, D. (2011). Quality Metrics to Accompany Spike Sorting of Extracellular Signals. *Journal of Neuroscience*, 31(24):8699–8705.

Hintiryan, H., Foster, N. N., Bowman, I., Bay, M., Song, M. Y., Gou, L., Yamashita, S., Bienkowski, M. S., Zingg, B., Zhu, M., Yang, X. W., Shih, J. C., Toga, A. W., and Dong, H.-W. (2016). The mouse cortico-striatal projectome. *Nature Neuroscience*, 19(8):1100–1114.

Hintzen, A., Pelzer, E. A., and Tittgemeyer, M. (2018). Thalamic interactions of cerebellum and basal ganglia. *Brain Structure and Function*, 223(2):569–587.

Hodos, W. (1961). Progressive Ratio as a Measure of Reward Strength. *Science*, 134(3483):943–944.

Hong, S. and Hikosaka, O. (2008). The globus pallidus sends reward-related signals to the lateral habenula. *Neuron*, 60(4):720–729.

Hooks, B. M., Papale, A. E., Paletzki, R. F., Feroze, M. W., Eastwood, B. S., Couey, J. J., Winnubst, J., Chandrashekar, J., and Gerfen, C. R. (2018). Topographic precision in sensory and motor corticostriatal projections varies across cell type and cortical area. *Nature Communications*, 9(1):3549.

Houk, J. C. and Wise, S. P. (1995). Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: Their role in planning and controlling action. *Cerebral cortex*, 5(2):95–110.

Hunnicutt, B. J., Jongbloets, B. C., Birdsong, W. T., Gertz, K. J., Zhong, H., and Mao, T. (2016). A comprehensive excitatory input map of the striatum reveals novel functional organization. *eLife*, 5:e19103.

Iakubovskii, P. (2019). Segmentation Models Pytorch.

Iigaya, K., Hauser, T. U., Kurth-Nelson, Z., O'Doherty, J. P., Dayan, P., and Dolan, R. J. (2020). The value of what's to come: Neural mechanisms coupling prediction error and the utility of anticipation. *Science Advances*, 6(25):eaba3828.

Inagaki, H. K., Chen, S., Ridder, M. C., Sah, P., Li, N., Yang, Z., Hasanbegovic, H., Gao, Z., Gerfen, C. R., and Svoboda, K. (2022). A midbrain-thalamus-cortex circuit reorganizes cortical dynamics to initiate movement. *Cell*, 185(6):1065–1081.e23.

Inagaki, H. K., Inagaki, M., Romani, S., and Svoboda, K. (2018). Low-Dimensional and Monotonic Preparatory Activity in Mouse Anterior Lateral Motor Cortex. *Journal of Neuroscience*, 38(17):4163–4185.

Iwamuro, H., Tachibana, Y., Ugawa, Y., Saito, N., and Nambu, A. (2017). Information processing from the motor cortices to the subthalamic nucleus and globus pallidus and their somatotopic organizations revealed electrophysiologically in monkeys. *European Journal of Neuroscience*, 46(11):2684–2701.

Jeljeli, M., Strazielle, C., Caston, J., and Lalonde, R. (2003). Effects of ventrolateral-ventromedial thalamic lesions on motor coordination and spatial orientation in rats. *Neuroscience Research*, 47(3):309–316.

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

Juavinett, A. L., Bekheet, G., and Churchland, A. K. (2018). Chronically-implanted Neuropixels probes enable high yield recordings in freely moving mice.

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydın, Ç., Barbic, M., Blanche, T. J., Bonin, V., Couto, J., Dutta, B., Gratiy, S. L., Gutnisky, D. A., Häusser, M., Karsh,

B., Ledochowitsch, P., Lopez, C. M., Mitelut, C., Musa, S., Okun, M., Pachitariu, M., Putzeys, J., Rich, P. D., Rossant, C., Sun, W.-l., Svoboda, K., Carandini, M., Harris, K. D., Koch, C., O'Keefe, J., and Harris, T. D. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232–236.

Jurkowski, A. J., Stepp, E., and Hackley, S. A. (2005). Variable foreperiod deficits in Parkinson's disease: Dissociation across reflexive and voluntary behaviors. *Brain and Cognition*, 58(1):49–61.

Kalkhoven, C., Sennef, C., Peeters, A., and van den Bos, R. (2014). Risk-taking and pathological gambling behavior in Huntington's disease. *Frontiers in Behavioral Neuroscience*, 8.

Kandel, E. R., editor (2013). *Principles of Neural Science*. McGraw-Hill, New York, 5th ed edition.

Kapoor, V., Besserve, M., Logothetis, N. K., and Panagiotaropoulos, T. I. (2018). Parallel and functionally segregated processing of task phase and conscious content in the prefrontal cortex. *Communications Biology*, 1(1):1–12.

Kawagoe, R., Takikawa, Y., and Hikosaka, O. (1998). Expectation of reward modulates cognitive signals in the basal ganglia. *Nature Neuroscience*, 1(5):411–416.

Kawai, R., Markman, T., Poddar, R., Ko, R., Fantana, A. L., Dhawale, A. K., Kampff, A. R., and Ölveczky, B. P. (2015). Motor cortex is required for learning but not for executing a motor skill. *Neuron*, 86(3):800–812.

Kelly, R. C., Smith, M. A., Samonds, J. M., Kohn, A., Bonds, A. B., Movshon, J. A., and Sing Lee, T. (2007). Comparison of Recordings from Microelectrode Arrays and Single Electrodes in the Visual Cortex. *Journal of Neuroscience*, 27(2):261–264.

Kha, H. T., Finkelstein, D. I., Tomas, D., Drago, J., Pow, D. V., and Horne, M. K. (2001). Projections from the substantia nigra pars reticulata to the motor thalamus of the rat: Single axon reconstructions and immunohistochemical study. *The Journal of Comparative Neurology*, 440(1):20–30.

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

Kimura, M. (1995). Role of basal ganglia in behavioral learning. *Neuroscience Research*, 22(4):353–358.

King, J.-R. and Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, 18(4):203–210.

Klein, P.-A., Olivier, E., and Duque, J. (2012). Influence of Reward on Corticospinal Excitability during Movement Preparation. *Journal of Neuroscience*, 32(50):18124–

18136.

Klockgether, T., Schwarz, M., Turski, L., and Sontag, K.-H. (1986). The rat ventromedial thalamic nucleus and motor control: Role of N-methyl-D-aspartate-mediated excitation, GABAergic inhibition, and muscarinic transmission. *Journal of Neuroscience*, 6(6):1702–1711.

Koren, V. (2021). Uncovering structured responses of neural populations recorded from macaque monkeys with linear support vector machines. *STAR Protocols*, 2(3):100746.

Kornhuber, H. H. and Deecke, L. (1965). [CHANGES IN THE BRAIN POTENTIAL IN VOLUNTARY MOVEMENTS AND PASSIVE MOVEMENTS IN MAN: READINESS POTENTIAL AND REAFFERENT POTENTIALS]. *Pflugers Archiv Fur Die Gesamte Physiologie Des Menschen Und Der Tiere*, 284:1–17.

Krantz, J. (2012). *Experiencing Sensation and Perception*. Pearson Education, Limited.

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

Kuramoto, E., Fujiyama, F., Nakamura, K. C., Tanaka, Y., Hioki, H., and Kaneko, T. (2011). Complementary distribution of glutamatergic cerebellar and GABAergic basal ganglia afferents to the rat motor thalamic nuclei. *European Journal of Neuroscience*, 33(1):95–109.

Kuramoto, E., Furuta, T., Nakamura, K. C., Unzai, T., Hioki, H., and Kaneko, T. (2009). Two Types of Thalamocortical Projections from the Motor Thalamic Nuclei of the Rat: A Single Neuron-Tracing Study Using Viral Vectors. *Cerebral Cortex*, 19(9):2065–2077.

Kuramoto, E., Ohno, S., Furuta, T., Unzai, T., Tanaka, Y. R., Hioki, H., and Kaneko, T. (2015). Ventral Medial Nucleus Neurons Send Thalamocortical Afferents More Widely and More Preferentially to Layer 1 than Neurons of the Ventral Anterior–Ventral Lateral Nuclear Complex in the Rat. *Cerebral Cortex*, 25(1):221–235.

Lanciego, J. L., Luquin, N., and Obeso, J. A. (2012). Functional Neuroanatomy of the Basal Ganglia. *Cold Spring Harbor Perspectives in Medicine*, 2(12):a009621.

Lau, B. and Glimcher, P. W. (2007). Action and Outcome Encoding in the Primate Caudate Nucleus. *Journal of Neuroscience*, 27(52):14502–14514.

Lau, B. and Glimcher, P. W. (2008). Value Representations in the Primate Striatum during Matching Behavior. *Neuron*, 58(3):451–463.

Lau, B., Monteiro, T., and Paton, J. J. (2017). The many worlds hypothesis of dopamine prediction error: Implications of a parallel circuit architecture in the basal ganglia. *Current Opinion in Neurobiology*, 46:241–247.

Lauwereyns, J., Watanabe, K., Coe, B., and Hikosaka, O. (2002). A neural correlate of response bias in monkey caudate nucleus. *Nature*, 418(6896):413–417.

Lee, D., Seo, H., and Jung, M. W. (2012). Neural Basis of Reinforcement Learning and Decision Making. *Annual Review of Neuroscience*, 35(1):287–308.

Levcik, D., Sugi, A. H., Pochapski, J. A., Baltazar, G., Pulido, L. N., Villas-Boas, C., Aguilar-Rivera, M., Fuentes-Flores, R., Nicola, S. M., and Cunha, C. D. (2021). Nucleus accumbens neurons encode initiation and vigor of reward approach behavior.

Li, Y., Lindemann, C., Goddard, M. J., and Hyland, B. I. (2016). Complex Multiplexing of Reward-Cue- and Licking-Movement-Related Activity in Single Midline Thalamus Neurons. *Journal of Neuroscience*, 36(12):3567–3578.

Libet, B., Gleason, C. A., Wright, E. W., and Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain: A Journal of Neurology*, 106 (Pt 3):623–642.

Lima, S. Q., Hromádka, T., Znamenskiy, P., and Zador, A. M. (2009). PINP: A New Method of Tagging Neuronal Populations for Identification during In Vivo Electrophysiological Recording. *PLOS ONE*, 4(7):e6099.

Lintz, M. J. and Felsen, G. (2016). Basal ganglia output reflects internally-specified movements. *eLife*, 5:e13833.

Lopes, G., Bonacchi, N., Frazão, J., Neto, J. P., Atallah, B. V., Soares, S., Moreira, L., Matias, S., Itskov, P. M., Correia, P.-c. A., Medina, R. E., Calcaterra, L., Dreosti, E., Paton, J. J., and Kampff, A. R. (2015). Bonsai: An event-based framework for processing and controlling data streams. *Frontiers in Neuroinformatics*, 9.

Macpherson, T., Matsumoto, M., Gomi, H., Morimoto, J., Uchibe, E., and Hikida, T. (2021). Parallel and hierarchical neural mechanisms for adaptive and predictive behavioral control. *Neural Networks*, 144:507–521.

Masset, P., Ott, T., Lak, A., Hirokawa, J., and Kepecs, A. (2020). Behavior- and Modality-General Representation of Confidence in Orbitofrontal Cortex. *Cell*, 182(1):112–126.e18.

Matsumoto, M. and Hikosaka, O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*, 447(7148):1111–1115.

Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Number v. 42 in Boston Studies in the Philosophy of Science. D. Reidel Pub. Co, Dordrecht, Holland ; Boston.

Maurin, Y., Banrezes, B., Menetrey, A., Mailly, P., and Deniau, J. M. (1999). Three-dimensional distribution of nigrostriatal neurons in the rat: Relation to the topog-

raphy of striatonigral projections. *Neuroscience*, 91(3):891–909.

McElvain, L. E., Chen, Y., Moore, J. D., Brigidi, G. S., Bloodgood, B. L., Lim, B. K., Costa, R. M., and Kleinfeld, D. (2021). Specific populations of basal ganglia output neurons target distinct brain stem areas while collateralizing throughout the diencephalon. *Neuron*, 109(10):1721–1738.e4.

McFarland, N. R. and Haber, S. N. (2000). Convergent Inputs from Thalamic Motor Nuclei and Frontal Cortical Areas to the Dorsal Striatum in the Primate. *The Journal of Neuroscience*, 20(10):3798–3813.

McFarland, N. R. and Haber, S. N. (2002). Thalamic Relay Nuclei of the Basal Ganglia Form Both Reciprocal and Nonreciprocal Cortical Connections, Linking Multiple Frontal Cortical Areas. *Journal of Neuroscience*, 22(18):8117–8132.

Mestre-Bach, G. and Potenza, M. N. (2023). Potential Biological Markers and Treatment Implications for Binge Eating Disorder and Behavioral Addictions. *Nutrients*, 15(4):827.

Milstein, D. and Dorris, M. (2011). The Relationship between Saccadic Choice and Reaction Times with Manipulations of Target Value. *Frontiers in Neuroscience*, 5.

Milstein, D. M. and Dorris, M. C. (2007). The Influence of Expected Value on Saccadic Preparation. *Journal of Neuroscience*, 27(18):4810–4818.

Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology*, 50(4):381–425.

Mink, J. W. and Thach, W. T. (1993). Basal ganglia intrinsic circuits and their role in behavior. *Current Opinion in Neurobiology*, 3(6):950–957.

Mioni, G., Capizzi, M., Vallesi, A., Correa, Á., Di Giacopo, R., and Stablum, F. (2018). Dissociating Explicit and Implicit Timing in Parkinson's Disease Patients: Evidence from Bisection and Foreperiod Tasks. *Frontiers in Human Neuroscience*, 12:17.

Miyachi, S. (2009). [Cortico-basal ganglia circuits–parallel closed loops and convergent/divergent connections]. *Brain and Nerve = Shinkei Kenkyu No Shinpo*, 61(4):351–359.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T. P., Harley, T., Silver, D., and Kavukcuoglu, K. (2016). Asynchronous Methods for Deep Reinforcement Learning. *arXiv:1602.01783 [cs]*.

Morris, R. (2008). Morris water maze. *Scholarpedia*, 3(8):6315.

Nakagawa, Y., Ishibashi, Y., Yoshii, T., and Tagashira, E. (1995). Muscimol induces state-dependent learning in Morris water maze task in rats. *Brain Research*, 681(1-2):126–130.

Nakahara, H. and Hikosaka, O. (2012). Learning to represent reward structure: A key to adapting to complex environments. *Neuroscience Research*, 74(3):177–183.

Nakahara, H., Nakamura, K., and Hikosaka, O. (2006). Extended LATER model can account for trial-by-trial variability of both pre- and post-processes. *Neural Networks*, 19(8):1027–1046.

Nambu, A. (2011). Somatotopic organization of the primate Basal Ganglia. *Frontiers in Neuroanatomy*, 5:26.

Nambu, A. and Tachibana, Y. (2014). Mechanism of parkinsonian neuronal oscillations in the primate basal ganglia: Some considerations based on our recent work. *Frontiers in Systems Neuroscience*, 8.

Nambu, A., Tokuno, H., and Takada, M. (2002). Functional significance of the cortico–subthalamo–pallidal 'hyperdirect' pathway. *Neuroscience Research*, 43(2):111–117.

Nauta, W. J. (1972). Neural associations of the frontal cortex. *Acta Neurobiologiae Experimentalis*, 32(2):125–140.

Nickerson, R. S., Collins, A. M., and Markowitz, J. (1969). Effects of uncertain warning signals on reaction time. *Perception & Psychophysics*, 5(2):107–112.

Nishimura, Y., Takada, M., and Mizuno, N. (1997). Topographic distribution and collateral projections of the two major populations of nigrothalamic neurons.: A retrograde labeling study in the rat. *Neuroscience Research*, 28(1):1–9.

Noorani, I. and Carpenter, R. (2016). The LATER model of reaction time and decision. *Neuroscience & Biobehavioral Reviews*, 64:229–251.

North, B. V., Curtis, D., and Sham, P. C. (2002). A Note on the Calculation of Empirical P Values from Monte Carlo Procedures. *American Journal of Human Genetics*, 71(2):439–441.

Nunez, J. (2008). Morris Water Maze Experiment. *Journal of Visualized Experiments : JoVE*, (19):897.

Okada, K., Hashimoto, K., and Kobayashi, K. (2022). Cholinergic regulation of object recognition memory. *Frontiers in Behavioral Neuroscience*, 16:996089.

Okoro, S. U., Goz, R. U., Njeri, B. W., Harish, M., Ruff, C. F., Ross, S. E., Gerfen, C., and Hooks, B. M. (2022). Organization of Cortical and Thalamic Input to Inhibitory Neurons in Mouse Motor Cortex. *The Journal of Neuroscience*, 42(43):8095–8112.

Oorschot, D. E. (1996). Total number of neurons in the neostriatal, pallidal, subthalamic, and substantia nigral nuclei of the rat basal ganglia: A stereological study using the cavalieri and optical disector methods. *The Journal of Comparative Neurology*, 366(4):580–599.

Orofino, a. G., Ruarte, M. B., and Alvarez, E. O. (1999). Exploratory behaviour after intra-accumbens histamine and/or histamine antagonists injection in the rat. *Behavioural brain research*, 102(1-2):171–80.

Osorio-Gómez, D., Guzmán-Ramos, K., and Bermúdez-Rattoni, F. (2022). Dopamine activity on the perceptual salience for recognition memory. *Frontiers in Behavioral Neuroscience*, 16:963739.

Oswal, A., Ogden, M., and Carpenter, R. (2007). The Time Course of Stimulus Expectation in a Saccadic Decision Task. *Journal of Neurophysiology*, 97(4):2722–2730.

Othman, M. Z., Hassan, Z., and Che Has, A. T. (2022). Morris water maze: A versatile and pertinent tool for assessing spatial learning and memory. *Experimental Animals*, 71(3):264–280.

Pallanti, S., Haznedar, M. M., Hollander, E., Licalzi, E. M., Bernardi, S., Newmark, R., and Buchsbaum, M. S. (2010). Basal Ganglia activity in pathological gambling: A fluorodeoxyglucose-positron emission tomography study. *Neuropsychobiology*, 62(2):132–138.

Panigrahi, B., Martin, K. A., Li, Y., Graves, A. R., Vollmer, A., Olson, L., Mensh, B. D., Karpova, A. Y., and Dudman, J. T. (2015). Dopamine Is Required for the Neural Representation and Control of Movement Vigor. *Cell*, 162(6):1418–1430.

Pardo-Vazquez, J. L., Castiñeiras-de Saa, J. R., Valente, M., Damião, I., Costa, T., Vicente, M. I., Mendonça, A. G., Mainen, Z. F., and Renart, A. (2019). The mechanistic foundation of Weber's law. *Nature Neuroscience*, 22(9):1493–1502.

Parent, A. and Hazrati, L.-N. (1995a). Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Research Reviews*, 20(1):91–127.

Parent, A. and Hazrati, L.-N. (1995b). Functional anatomy of the basal ganglia. II. The place of subthalamic nucleus and external pallidium in basal ganglia circuitry. *Brain Research Reviews*, 20(1):128–154.

Park, J., Coddington, L. T., and Dudman, J. T. (2020). Basal Ganglia Circuits for Action Specification. *Annual Review of Neuroscience*, 43(1):485–507.

Paxinos, G. and Watson, C. (1998). *The Rat Brain in Stereotaxic Coordinates*. Academic Press, San Diego, 4th ed edition.

Paxinos, G. and Watson, C. (2009). *The Rat Brain in Stereotaxic Coordinates*. Elsevier Academic Press, Amsterdam Heidelberg, compact 6. ed edition.

Rapoport, J. L. (1990). Obsessive compulsive disorder and basal ganglia dysfunction. *Psychological Medicine*, 20(3):465–469.

Reger, M. L., Hovda, D. A., and Giza, C. C. (2009). Ontogeny of Rat Recognition Memory Measured by the Novel Object Recognition Task. *Developmental psychobi-*

ology, 51(8):672–678.

Rizzi, G. and Tan, K. R. (2019). Synergistic Nigral Output Pathways Shape Movement. *Cell Reports*, 27(7):2184–2198.e4.

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

Rudebeck, P. H. and Murray, E. A. (2014). The orbitofrontal oracle: Cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. *Neuron*, 84(6):1143–1156.

Rudell, A. P. and Hu, B. (2001). Does a warning signal accelerate the processing of sensory information? Evidence from recognition potential responses to high and low frequency words. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 41(1):31–42.

Rushworth, M. F. S., Noonan, M. P., Boorman, E. D., Walton, M. E., and Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron*, 70(6):1054–1069.

Saez, A., Rigotti, M., Ostojic, S., Fusi, S., and Salzman, C. D. (2015). Abstract Context Representations in Primate Amygdala and Prefrontal Cortex. *Neuron*, 87(4):869–881.

Sakagami, T. and Lattal, K. A. (2016). The Other Shoe: An Early Operant Conditioning Chamber for Pigeons. *The Behavior Analyst*, 39(1):25–39.

Sakai, S. T., Grofova, I., and Bruce, K. (1998). Nigrothalamic projections and nigrothalamocortical pathway to the medial agranular cortex in the rat: Single- and double-labeling light and electron microscopic studies. *The Journal of Comparative Neurology*, 391(4):506–525.

Sato, M. and Hikosaka, O. (2002). Role of Primate Substantia Nigra Pars Reticulata in Reward-Oriented Saccadic Eye Movement. *Journal of Neuroscience*, 22(6):2363–2373.

Schäfer, C. B., Gao, Z., De Zeeuw, C. I., and Hoebeek, F. E. (2021). Temporal dynamics of the cerebello-cortical convergence in ventro-lateral motor thalamus. *The Journal of Physiology*, 599(7):2055–2073.

Schultz, W. (1986). Activity of pars reticulata neurons of monkey substantia nigra in relation to motor, sensory, and complex events. *Journal of Neurophysiology*, 55(4):660–677.

Schultz, W., Apicella, P., and Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *The Journal of neuroscience : the official journal of the*

Society for Neuroscience, 13(3):900–13.

Schurger, A., Hu, P. B., Pak, J., and Roskies, A. L. (2021). What Is the Readiness Potential? *Trends in Cognitive Sciences*, 25(7):558–570.

Schweimer, J. and Hauber, W. (2005). Involvement of the rat anterior cingulate cortex in control of instrumental responses guided by reward expectancy. *Learning & memory (Cold Spring Harbor, N.Y.)*, 12(3):334–42.

Sclafani, A. and Ackroff, K. (2003). Reinforcement value of sucrose measured by progressive ratio operant licking in the rat. *Physiology & Behavior*, 79(4–5):663–670.

Scott, B. B., Constantinople, C. M., Akrami, A., Hanks, T. D., Brody, C. D., and Tank, D. W. (2017). Fronto-parietal Cortical Circuits Encode Accumulated Evidence with a Diversity of Timescales. *Neuron*, 95(2):385–398.e5.

Seger, C. A. and Spiering, B. J. (2011). A Critical Review of Habit Learning and the Basal Ganglia. *Frontiers in Systems Neuroscience*, 5.

Shadmehr, R. (2020). *Vigor: Neuroeconomics of Movement Control*. The MIT Press, Cambridge, Massachusetts.

Shadmehr, R., Huang, H. J., and Ahmed, A. A. (2016). Effort, reward, and vigor in decision-making and motor control. *Current biology : CB*, 26(14):1929–1934.

Shadmehr, R., Reppert, T. R., Summerside, E. M., Yoon, T., and Ahmed, A. A. (2019). Movement Vigor as a Reflection of Subjective Economic Utility. *Trends in Neurosciences*, 42(5):323–336.

Sharot, T. (2011). The optimism bias. *Current Biology*, 21(23):R941–R945.

Shen, W., Flajolet, M., Greengard, P., and Surmeier, D. J. (2008). Dichotomous Dopaminergic Control of Striatal Synaptic Plasticity. *Science*, 321(5890):848–851.

Shenoy, K. V., Sahani, M., and Churchland, M. M. (2013). Cortical Control of Arm Movements: A Dynamical Systems Perspective. *Annual Review of Neuroscience*, 36(1):337–359.

Shevtsova, Z., Malik, J. M. I., Michel, U., Bähr, M., and Kügler, S. (2005). Promoters and serotypes: Targeting of adeno-associated virus vectors for gene transfer in the rat central nervous system in vitro and in vivo. *Experimental Physiology*, 90(1):53–59.

Simonyan, K. (2019). Recent advances in understanding the role of the basal ganglia. *F1000Research*, 8:F1000 Faculty Rev–122.

Skinner, B. F. (1938). *The Behavior of Organisms: An Experimental Analysis*. The Behavior of Organisms: An Experimental Analysis. Appleton-Century, Oxford, England.

Skinner, B. F. (1948). 'Superstition' in the pigeon. *Journal of Experimental Psychology*, 38(2):168–172.

Staddon, J. E. R. and Cerutti, D. T. (2003). Operant Conditioning. *Annual Review of Psychology*, 54(1):115–144.

Stephenson-Jones, M., Samuelsson, E., Ericsson, J., Robertson, B., and Grillner, S. (2011). Evolutionary Conservation of the Basal Ganglia as a Common Vertebrate Mechanism for Action Selection. *Current Biology*, 21(13):1081–1091.

Stephenson-Jones, M., Yu, K., Ahrens, S., Tucciarone, J. M., van Huijstee, A. N., Mejia, L. A., Penzo, M. A., Tai, L.-H., Wilbrecht, L., and Li, B. (2016). A basal ganglia circuit for evaluating action outcomes. *Nature*, 539(7628):289–293.

Steverson, K., Chung, H.-K., Zimmermann, J., Louie, K., and Glimcher, P. (2019). Sensitivity of reaction time to the magnitude of rewards reveals the cost-structure of time. *Scientific Reports*, 9(1):20053.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, Mass.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning Series. The MIT Press, Cambridge, MA, second edition edition.

Swets, J. A. (1961). Detection theory and psychophysics: A review. *Psychometrika*, 26:49–63.

Sych, Y., Fomins, A., Novelli, L., and Helmchen, F. (2022). Dynamic reorganization of the cortico-basal ganglia-thalamo-cortical network during task learning. *Cell Reports*, 40(12):111394.

Takikawa, Y., Kawagoe, R., Itoh, H., Nakahara, H., and Hikosaka, O. (2002). Modulation of saccadic eye movements by predicted reward outcome. *Experimental Brain Research*, 142(2):284–291.

Tecuapetla, F., Jin, X., Lima, S. Q., and Costa, R. M. (2016). Complementary Contributions of Striatal Projection Pathways to Action Initiation and Execution. *Cell*, 166(3):703–715.

Thompson, T. (1968). Drugs as Reinforcers: Experimental Addiction. *International Journal of the Addictions*, 3(1):199–206.

Thorn, C. A., Atallah, H., Howe, M., and Graybiel, A. M. (2010). Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron*, 66(5):781–795.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, 2(4):i–109.

Thura, D. and Cisek, P. (2014). Deliberation and commitment in the premotor and primary motor cortex during dynamic decision making. *Neuron*, 81(6):1401–1416.

Tobler, P. N., Fiorillo, C. D., and Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. *Science (New York, N.Y.)*, 307(5715):1642–5.

Tootell, R. B., Switkes, E., Silverman, M. S., and Hamilton, S. L. (1988). Functional anatomy of macaque striate cortex. II. Retinotopic organization. *Journal of Neuroscience*, 8(5):1531–1568.

Upper, D. (1974). The unsuccessful self-treatment of a case of "writer's block". *Journal of Applied Behavior Analysis*, 7(3):497.

Valjent, E. and Gangarossa, G. (2021). The Tail of the Striatum: From Anatomy to Connectivity and Function. *Trends in Neurosciences*, 44(3):203–214.

Vallesi, A., Arbula, S., and Bernardis, P. (2014). Functional dissociations in temporal preparation: Evidence from dual-task performance. *Cognition*, 130(2):141–151.

van Daal, R. J. J., Aydin, Ç., Michon, F., Aarts, A. A. A., Kraft, M., Kloosterman, F., and Haesler, S. (2021). Implantation of Neuropixels probes for chronic recording of neuronal activity in freely behaving mice and rats. *Nature Protocols*, 16(7):3322–3347.

van Dijk, K. J., Janssen, M. L. F., Zwartjes, D. G. M., Temel, Y., Visser-Vandewalle, V., Veltink, P. H., Benazzouz, A., and Heida, T. (2016). Spatial Localization of Sources in the Rat Subthalamic Motor Region Using an Inverse Current Source Density Method. *Frontiers in Neural Circuits*, 10.

Verharen, J. P. H., Adan, R. A. H., and Vanderschuren, L. J. M. J. (2019). Differential contributions of striatal dopamine D1 and D2 receptors to component processes of value-based decision making. *Neuropsychopharmacology*, 44(13):2195–2204.

von Uexküll, J. and von Uexküll, J. (2010). *A Foray into the Worlds of Animals and Humans: With A Theory of Meaning*. Number 12 in Posthumanities. University of Minnesota Press, Minneapolis, 1st university of minnesota press ed edition.

Vyas, S., Golub, M. D., Sussillo, D., and Shenoy, K. V. (2020). Computation Through Neural Population Dynamics. *Annual Review of Neuroscience*, 43(1):249–275.

Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., Hassabis, D., and Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience*, 21(6):860–868.

Wang, Y., Yin, X., Zhang, Z., Li, J., Zhao, W., and Guo, Z. V. (2021). A cortico-basal ganglia-thalamo-cortical channel underlying short-term memory. *Neuron*, page S0896627321005778.

Wei, W. and Wang, X.-J. (2016). Inhibitory Control in the Cortico-Basal Ganglia-Thalamocortical Loop: Complex Regulation and Interplay with Memory and Deci-

sion Processes. *Neuron*, 92(5):1093–1105.

Welford, A. (1986). Note on the Effects of Practice on Reaction Times. *Journal of Motor Behavior*, 18(3):343–345.

Welker, C. (1971). Microelectrode delineation of fine grain somatotopic organization of (SmI) cerebral neocortex in albino rat. *Brain Research*, 26(2):259–275.

Wise, R. A. (2009). Roles for nigrostriatal–not just mesocorticolimbic–dopamine in reward and addiction. *Trends in neurosciences*, 32(10):517–24.

Wise, R. A. and Bozarth, M. A. (1982). Action of drugs of abuse on brain reward systems: An update with specific attention to opiates. *Pharmacology Biochemistry and Behavior*, 17(2):239–243.

Wise, S. P. (1996). The role of the basal ganglia in procedural memory. *Seminars in Neuroscience*, 8(1):39–46.

Wittgenstein, L. and Ogden, C. K. (1999). *Tractatus Logico-Philosophicus*. Dover Publications, Mineola, N Y.

Woolsey, C. N., Settlage, P. H., Meyer, D. R., Sencer, W., Pinto Hamuy, T., and Travis, A. M. (1952). Patterns of localization in precentral and "supplementary" motor areas and their relation to the concept of a premotor area. *Research Publications - Association for Research in Nervous and Mental Disease*, 30:238–264.

Worden, R., Bennett, M. S., and Neacsu, V. (2021). The Thalamus as a Blackboard for Perception and Planning. *Frontiers in Behavioral Neuroscience*, 15:27.

Xiao, D., Zikopoulos, B., and Barbas, H. (2009). Laminar and modular organization of prefrontal projections to multiple thalamic nuclei. *Neuroscience*, 161(4):1067–1081.

Yang, W., Tipparaju, S. L., Chen, G., and Li, N. (2022). Thalamus-driven functional populations in frontal cortex support decision-making. *Nature Neuroscience*, 25(10):1339–1352.

Yasuda, M. and Hikosaka, O. (2015). Functional territories in primate substantia nigra pars reticulata separately signaling stable and flexible values. *Journal of Neurophysiology*, 113(6):1681–1696.

Yasuda, M. and Hikosaka, O. (2018). Medial thalamus in the territory of oculomotor basal ganglia represents stable object value. *European Journal of Neuroscience*, 0(0).

Yin, H. H., Mulcare, S. P., Hilário, M. R. F., Clouse, E., Holloway, T., Davis, M. I., Hansson, A. C., Lovinger, D. M., and Costa, R. M. (2009). Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nature Neuroscience*, 12(3):333–341.

Yoon, T., Geary, R. B., Ahmed, A. A., and Shadmehr, R. (2018). Control of movement vigor and decision making during foraging. *Proceedings of the National Academy of*

Sciences, 115(44):E10476–E10485.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation.

ITQB NOVA, Avenida da República, 2780-157 Oeiras, Portugal | Tel +351 214 469 100 | itqb.unl.pt