Computational modeling of cognitive control for rule-guided behavior

Snigdha Dagar

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR
DE L'UNIVERSITÉ DE BORDEAUX

ECOLE DOCTORALE : Mathématiques et Informatique

SPÉCIALITÉ : Informatique

Par **Snigdha DAGAR**

Computational modeling of Cognitive Control for rule-guided behavior

Sous la direction de : **Frédéric ALEXANDRE**
Co-directeur : **Nicolas P. ROUGIER**

Soutenue le 21 Avril 2023

Membres du jury :

Mme. Lola CANAMERO	Professeur	CY Cergy Paris Université	Rapportrice
M. Emmanuel PROCYK	Directeur de Recherche	Inserm	Rapporteur
M. Julien VITAY	Chargé de Recherche	Chemnitz University of Technology	Examinateur
Mme Macha NIKOLSKI	Directrice de recherche	Université de Bordeaux	Présidente
M. Frédéric ALEXANDRE	Directeur de recherche	Université de Bordeaux	Directeur
M. Nicolas P. ROUGIER	Directeur de recherche	Université de Bordeaux	Directeur

Modélisation du contrôle cognitif pour le comportement guidé par les règles

Résumé : Le contrôle cognitif est la capacité générale d'un organisme à inhiber le comportement dominant en faveur d'une réponse pertinente selon des objectifs internes et en lien avec des facteurs environnementaux et/ou motivationnels. Diverses études expérimentales ainsi que des modèles computationnels ont tenté de mettre en évidence les mécanismes et les structures neuronales sous-jacents qui autorisent un comportement à la fois flexible et adaptatif. Néanmoins, une théorie unifiée qui tiendrait compte de l'ensemble de ces mécanismes reste insaisissable, notamment en ce qui concerne le degré d'adaptabilité qui varie entre les humains et les animaux non humains. Dans ce travail, nous souhaitons caractériser cette gradation du contrôle cognitif afin de poser un cadre conceptuel nous permettant de concevoir des modèles informatiques biologiquement plausibles à même de mettre en évidence les étapes clés du contrôle cognitif.

Dans une première approche, et sur la base d'études chez la souris, nous utilisons un modèle acteur-critique standard afin de montrer comment le comportement naturel d'exploration de la souris doit être inhibé afin de permettre au modèle d'apprendre une règle simple dans un labyrinthe radial. Au travers d'une série de tâche de complexité croissante, nous montrons alors la nécessité de posséder des systèmes de mémoire de travail et épisodique, en adéquation avec la littérature sur la prise de décision chez les rongeurs. Cela est notamment réalisé en étendant le modèle précédent avec une modélisation fonctionnelle de ces deux systèmes de mémoire, nous permettant ainsi de caractériser les contributions respectives de ces deux systèmes, en accord avec les études chez les rongeurs.

Dans un troisième temps, nous mettons en évidence la nécessité de former des représentations explicites du contexte à partir de règles acquises implicitement, et ceci, afin de pouvoir acquérir un comportement spécifique vis à vis d'un contexte particulier. Enfin, pour comprendre comment le cortex préfrontal soutient cet apprentissage contextuel et autorise une pleine capacité du contrôle cognitif chez l'Homme, nous proposons un modèle hiérarchique global qui explique notamment le rôle de l'attention sélective dans l'apprentissage de règles abstraites. Notre hypothèse étant que cette capacité d'attention permet la sélection des règles concrètes les plus appropriées ainsi que la manipulation des représentations sous-jacentes. Tout cela étant réalisé en assurant le monitoring des ces représentations ainsi que les erreurs de prédiction.

Mots-clés : Contrôle cognitif, Neurosciences computationnelles, apprentissage par renforcement, Cortex Préfrontal

---

Computational modeling of Cognitive Control for Rule-guided behavior

Abstract: Cognitive Control is the general capacity of an organism to use top-down control signals to inhibit the dominant behavior in favor of a contextually relevant response, in accordance with internally described goals (which could result from environmental or motivational factors).

Various experimental studies and computational models have tried to understand the neural mechanisms and structures that enable flexible and adaptive behavior by exerting cognitive control. Nevertheless, a unifying theory that explains these mechanisms remains elusive. The degree of adaptability that cognitive control provides varies from humans to nonhuman animals. We elaborate this gradation of cognitive control in a conceptual framework, and then use biologically plausible computational models to identify key computational processing requirements at each stage.

In the first model, we use a basic actor-critic model, to show how the default behavior of exploration in mice, needs to be overridden in order for a rodent (agent) to learn a simple tactile rule in a radial maze. Based on the decision making literature on rodents, we then show through a series of incrementally complex tasks, the necessity of working and episodic memory systems. This is done by extending the previous model with an elementary abstraction of these memory systems in order to make concrete the underlying mechanisms and criteria of cognitive control in rodents. As a third step, we highlight the need to form explicit mental representations of "context" from implicitly acquired rules, to enable contextually guided behavior, using a simple recurrent neural network trained on a sensorimotor task. Finally, to understand how the PFC supports contextual learning and the full capacity of cognitive control in humans, we develop a hierarchical computational model that explains the role of selective and sustained attention in learning abstract rules, and selects the appropriate concrete rules by manipulating representations, or task sets, and monitoring these representations and prediction errors.

Keywords: Cognitive Control, Computational Neuroscience, Reinforcement learning, Prefrontal Cortex

---

Unité de recherche
LaBRI, UMR 5800, Université, 33000 Bordeaux, France.

Acknowledgements

I owe my deepest gratitude to my thesis supervisors, Fréderic Alexandre, and Nicolas Rougier, for their guidance over the last three years. The lessons I have learnt from you are too many to list, but first and foremost, you have taught me that the research journey is about asking the right questions, to look at both the big picture, and the minutest details. You have let me arrive at them in my own time and pace, and encouraged me to think for myself. No discussion with you has ever left me feeling less than inspired, and invigorated about this incredibly interesting topic that has become my thesis. Thank you for your patience, your detailed feedback and your constant encouragement. More than anything, thank you for believing in me. It has been a pleasure and honor to do research under your wings.

My work would not have been complete without the collaboration with Christopher. Discussions with you have been very influential in my understanding of experimental behavioral work, and has sharpened my sense of where we, as computational modelers fit in, and what we are seeking to explain. Special thanks to Naomi - research as a PhD student can be lonely, and co-working with you was inspiring, and motivated me to resolve the nitty gritty details of the computational process. I extend my thanks also to Nathan and Xavier, who helped us with digging into the "reservoir" with their expertise.

I would also like to acknowledge the contribution of the "Controlled Contextual Learning" group in my work - Thierry and Jianyong who led the group and initiated excellent scientific discussions. Thanks to Hugo, with who I was able to discuss the larger perspective of PFC-hippocampus interactions, and Chloé, for the symbolic processing perspective.

Although not directly related to my research, there are a number of people without who, this journey would not have been possible, and to who I will always be indebted.

I would like to extend my sincerest gratitude to the former team of Mnemosyne - Anthony, Thalita, Pramod, Ikram, Bhargav, and beloved room mate Silvia - who made me feel so at home in this city from the first day, and throughout the lockdown days of covid. We didn't get to spend much time in the same city, and yet, your absence was always felt.

To the entire team of Mnemosyne and Neuroprosthetics - Nikos, Melodie, Ankur, Axel. The years have passed by in a blink in between the coffee breaks, the memes, and the laughs.

To my favorite colocs - Giuliano and Fjola - you both made a "home". My sanity was preserved because of our daily dinners, our evening chats. Thank you for being my companions, for celebrating my small wins, and lifting me up when I was down.

To Remya, my constant over a decade, through all the ups and downs. You have made every hurdle simpler, a little easier to cross. You are my first reach out in times of trouble, that is suffice to say.

To my friends in London - Yash, Saumya, Garg (and Drushti in Amsterdam), thank you for keeping your doors open for me always, for being my cheerleaders and my much needed respite. Nam, Simran, Jigar, and Arwa - thank you for pampering me, I will find my way to your couch again. You have made a new meaning for home. My friends across the seas - you know who you are, but specially Arnav and Sudhanshu - thank you for your support.

Finally, to my one rock solid support system, my north stars - Siddharth, Vrinda. Mom, Dad - everything I do, I owe to you.

Contents

Introduction

The science of AI is concerned with the study of intelligent forms of behavior, in computational terms. But even to define the term intelligence is notoriously difficult. Is AI able to tell us when a good semblance of behavior can be achieved using cheap tricks that seem to have little to do with what we intuitively imagine intelligence to be ? Take ChatGPT for example. That it can automatically produce what me might consider to be humanlike text is remarkable. But roughly speaking, such AIs take huge amounts of data, search for patterns in it and become increasingly proficient at generating statistically probable output. So then, are these intuitions wrong, and is intelligence really just a bag of tricks ? Or are the philosophers right, and is a behavioral understanding of intelligence simply too weak ? [126]

To begin asking specific questions about behavior, we need to first define what sort of behavior we care about. Different researchers will quite naturally focus on different aspects. The behavior may or may not depend on perceptual or motor skills. It may or may not include learning. It may or may not be grounded in emotional responses, or in social interactions. From the perspective of neuroscience, if the question is "How does the brain lead to behavior", the pertinent question to first ask is *why* is the brain performing this behavior and then asking *how* is it doing it [117]. Moreover, understanding something is not the same as just describing it or knowing how to intervene to change it - an aspect on which AI and neuroscience have significantly diverged.

Consider the decisions we make on a daily basis about how to get around. Our preferred mode of transport is based on multiple individual, social and situational factors. Seemingly on daily recurrent journeys, it is a behavior we rarely intentionally think about. Depending on how one has grown up, the city one lives in, and how far one has to travel, we usually default to certain habitual patterns, or past repetitive behaviors. Yet, certain external or internal levers seem to be able to modify this choice, leading to deliberation and possibly the choice of an alternate behavior. These contextual factors could be available infrastructure (maybe there is a strike), weather conditions (biking or walking on a rainy day is unappealing),

time constraints (you have an appointment and are running on a tight schedule), or even internal motivational factors (you want to hit your target of 10,000 steps a day) [1]. Decision making is rarely ever as simple as choosing between two options. In the complex environment we live in, amending our behaviors is an explicit, deliberative, voluntary process that requires Cognitive Control.

For the purpose of this manuscript, it is this kind of *adaptive behavior* that is the spotlight of our study, or more precisely Cognitive Control over rule guided behavior. Many cognitive abilities in human and non human primates, such as inferential reasoning, planning, social interaction and flexibility in adapting to novel situations are known to be strongly dependent on the formation and implementation of rules, and impairments in such cognitive processes have been reported in a range of neuropsychological disorders.

There has been much focus on systems neuroscience in the past few decades, with the advent of advanced recording techniques pushing the envelope on the microscopic scale at which we can record neural data. It remains the case though that it is very hard to infer the mapping between the behavior of a system and its lower-level properties by only looking at the lower level properties. To quote Marr on the inadequacy of a strictly neurophysiological approach to understanding :
trying to understand perception by understanding neurons is like trying to understand a bird's flight by studying only feathers. It just cannot be done

In other words, understanding what process X is - what computations it embodies - can rarely be done at the neural level alone. Even when looking at neural data alone, some authors argue [173] that a mathematical bias by researchers may actually hinder our understanding of brain and cognition, the interpretations of said data, or even what data to look for in the first place. Using the impact of mathematics in the field of decision making and as explained in [127], they quote an example :
nearly all theories of decision, from expected utility theory through prospect theory and even modern reinforcement learning algorithms have shared the notion that in order to choose, the different attributes of each option must at some point be converged, however idiosyncratically, incompletely and imperfectly, into a single value for the actual process of comparison. They argue that this notion of value as the common currency equated to decisions may be a mathematical bias since mathematics offers no other way of comparing two unrelated objects.

On the behavioral science side, many have argued that examining behavior itself is more valuable that looking at individual neurons or neuronal assemblies [147].

[1] This PhD was part of a collaborative project "EcoMob", with other teams specialized in data science and economic choice, around the observation of choices made by human subjects in realistic multi-criteria tasks related to transportation habits in an eco-responsibility context

This idea is explained well through an example: A rat is trained to run through a T-shaped maze to reach food, and at the junction of the T, it must decide whether to turn right or left. The rat may use either internal cues (turning right relative to its own body) or external cues (turning towards a specific location in space) to make this decision. However, it is quite a challenge to determine which strategy the rat is using by recording brain activity alone. To investigate this behavior, Packard and McGaugh (1996) [153] turned the maze around and observed the rat's behavior. If the rat continued to turn right, it would indicate an egocentric strategy, but if it turned left, it must be following external cues. This manipulation allowed for a better understanding of the neural strategies used by the rat and led to follow-up experiments to determine the conditions under which rodents use allocentric rather than egocentric strategies. Ultimately, this research led to computational models that specified the types of data that the rat must learn and use in each condition, and the computations that may support transitioning from one strategy to another.

In the field of computational modeling of cognitive science, the discourse has moved from the connectionist framework that previously held much promise as an explanatory model of the brain, to a bayesian modeling of cognition [103]. Most proposals is these frameworks start from the same computational principle, that a goal function must be optimized. This may be an error function (to be minimized) in the PDP tradition, a value function (to be maximized) in the RL tradition, or a posterior probability function (to be maximized) in the bayesian tradition. Examining the claims of deep learning models of neuroscience, Schaeffer and colleagues in [182] have demonstrated that the results of these models are more strongly driven by *particular, non-fundamental, and post-hoc implementation choices than fundamental truths about neural circuits or the loss function(s) they might optimize*. These authors conclude that such models thus cannot be expected to produce accurate models of the brain without the addition of substantial amounts of inductive bias, and as such don't hold any explanatory sense for neuroscience.

Clearly, in trying to ask the right questions to investigate what mechanisms lead to adaptive behavior, one needs to traverse this rich and complex academic landscape, with regards to what normative approach to use, what assumptions to make, what modeling tradition to follow, what level of description is most useful etc. This manuscript is hence divided in two parts : Part I lays out the groundwork and the current understanding of Cognitive Control from the perspective of cognitive science, neuroscience, and computer science.

In chapter 1, we begin by presenting a conceptual overview that serves as the foundation for the rest of this manuscript. We explain classical behavioral theo-

ries that have formed our understanding of learning, and subsequently we motivate the need for Cognitive Control. The second half of the chapter then invokes conceptual notions associated with Cognitive Control - in terms of constructs, rules, context, abstraction, hierarchy, and in terms of mechanisms, working memory, attention and performance monitoring. Agnostic of the species, we present a conceptual view of Cognitive Control as a gradient, with different kinds of tasks requiring different degrees of control.

In chapter 2, we then provide the neuroscience evidence for the mechanisms involved, and discuss how differences in the cortex anatomy of different species can underlie the difference in complexity of cognitive control observed in their behaviors. Further, we gather evidence on the functional subdivisions in the primate prefrontal cortex and the roles they might subserve. We put forth a functional sketch of how these different brain regions interact with one another in learning, and cognitive control.

In chapter 3, we review computational models of cognitive control, analysing to what extent these existing models are able to explain and integrate the computational principles and biological mechanisms identified with cognitive control. At the end of this chapter, we identify some open questions with regard to the synthesis of the literature reviewed in the first three chapters.

Part II then details our particular contribution to the story of Cognitive Control, using different kinds of computational models. In chapter 4, we present a series of computational models to identify the roles of the working and episodic memory learning systems, default behavior and context, in the kind of cognitive control at work in rodents.

Finally, in chapter 5, we present a hierarchical model, in which superior layers in the hierarchy deploy selective attention to select among multiple possible options of responses at the lower layer.

We conclude this manuscript by a summary of our main findings and an introduction to some fascinating perspectives it has opened.

Part I

Cognitive Control

A comprehensive literature review

1 Conceptual overview

1.1 Introduction

In 1944, Skinner wrote the "Principles of Behavior" to present in an objective, systematic manner the primary, or fundamental, molar principles of behavior, with the assumption that all behavior, individual and social, moral and immoral, normal and psychopathic, is generated from the same primary laws [191]. While we have come a long way in the study of behavior since the publication of that book, the motivation remains the same and a long line of studies continue to deepen the work of unraveling the principles of behavior and cognition.

In that regards, a central and ubiquitous element of behavior is the capacity to make choices, that is, a commitment to a proposition among alternatives that arises through a process of deliberation. A series of decisions makes a pattern of behavior and consequently, the inner mechanisms of decision making provide a window on cognition. Many cognitive abilities in human and non human primates, such as inferential reasoning, planning, social interaction and flexibility in adapting to novel situations are contingent on complex decision making. These kind of executive functions are higher level functions for the control of cognition. Over the last few decades, scholars from economics, psychology, cognitive science, evolutionary biology, computer science and neuroscience have tried to provide a framework for investigating the neural mechanism of choice behavior, poised to investigate decision making at the theoretical, algorithmic and implementation levels [134].

This chapter aims at gathering information towards an important and difficult question : How does the brain choose efficiently and adaptively among available options to ensure coherent, goal-directed behavior ? Hidden behind this question are many problems that necessitate a multidisciplinary approach. Indeed, to understand how humans and other animals solve this problem, we need answers from researchers versed in anatomy, traditional psychology, learning theory, neuroimaging and mathematical modeling. Central to nearly all definitions of exec-

utive function are two concepts : *rules* and *control*. The rules that guide human behavior - and the behavior of many other but not all organisms are abstract and flexible. Control processes allow us to engage rules appropriate to a particular context. These two aspects — (1) creating and modifying rules for behavior and (2) engaging the appropriate rule for a particular context—represent the highest levels of a taxonomy for executive function and thus reflect major sections of this chapter.

1.2 Theories of behavior and learning

In the formalism of reinforcement learning, any kind of decision making rule or strategy consists of "a mapping from perceived states of the environment to actions to be taken when in those states" [195]. Historically, how reinforcement learning conceptualizes this mapping, is rooted in experimental and animal psychology. The study of animal conditioning is broadly divided into two main areas : Classical or Pavlovian and instrumental or operant conditioning.

First, described by Ivan Pavlov, Classical, Pavlovian or sensory (respondent) conditioning focuses on involuntary, automatic behaviors. It is a process that involves pairing a previously neutral (conditioned) stimulus (e.g., a bell) with an appetitive unconditioned (or biologically significant) stimulus (e.g., food); after repeated exposure to the food following the bell, the bell starts to elicit reflexive responses normally reserved for the food (e.g., salivation). It is the expression of innate knowledge (in other words, the decision making structures presumably evolved to handle natural rewards (*reinforcers*) such as food, water, and intrinsic threats) of what responses are usually appropriate when certain types of events are observed to be correlated in the environment. In the pavlovian scheme, responses are thus stereotyped (also called pavlovian reflexes) and are consequences of learned associations. The Rescorla-Wagner (RW) [168] model, arguably the most influential model of animal learning to date, provided the mathematical formalism that could capture this phenomenon by postulating that learning occurs *only when events violate expectations*. For instance, in a conditioning trial in which *conditional stimuli* CS_1 and CS_2 (say a light and a tone) were presented, as well as an affective stimulus (food), the *unconditional stimulus* US, the model postulates that the associative strength of each of the conditioned stimuli $V(CS_i)$ will change according to

$$V_{new}(CS_i) = V_{old}(CS_i) + \eta(CS_i, US) \times [\lambda(US) - \sum_i V_{old}(CS_i)]$$

Whereas, based on CS, pavlovian conditioning *passively* predicts the US to occur and prepares the body for this event, another more *active* learning scheme called Operant or instrumental conditioning focuses on using reinforcement (positive or negative) to increase or decrease the strength of a *voluntary* behavior. Notably, Thorndike (1911) first presented the concept of Law of Effect, which is the idea that a response will be triggered more (or less) frequently when observed to lead to a positive (or negative) consequence. Alternatively, a response will be triggered more frequently if it leads to the avoidance or removal of a negative stimulus. The early work of behaviorist experimentalist and theorist B.F. Skinner used the terminology *positive reinforcers* and *negative reinforcers* to describe these two. Computationally, this kind of learning is treated as attempting to optimize the consequences of actions in terms of some long term measure of total obtained rewards (and/or avoid punishments).

Since here, the response is voluntary, it is possible to consider the corresponding level of need, devaluate the outcome and refrain from acting if the motivation is low. Whereas pavlovian conditioning simply defines how much an outcome is liked (thus described as Stimulus-Response associations), instrumental conditioning considers how much it is currently wanted and chooses to trigger responses taking motivations into account (hence described as Response-Outcome associations). These can be extrinsic motivations, to get a desired (external or extrinsic) outcome satisfying fundamental needs, or more complex internal representations, expressing intrinsic motivations, as described in more detail in [152]. They are related to a more abstract need of (intrinsic) information, to obtain from the exploration of the complex world and from the monitoring of internal activity, as is the case with curiosity and attention towards novelty.

This categorization of behavior was motivated by the observable *responses* an animal made or could be made to exhibit through experimentation. On the other hand, another categorization of behavior has been motivated by the driving factors of behavior, i.e. about the internal or external environmental triggers for particular behaviors. Hence, external and stimulus-driven on the one hand or internal and goal-directed on the other, where a *goal* is any mental representation of a desired activity (which need not be a biologically significant stimulus). Instrumental conditioning can be performed under the control of (or conditional to) stimuli also called occasion setters, that can become conditional reinforcers, leading to chaining in complex behavioral goal-directed sequences towards primary reinforcers (respectively defined as subgoals and goals in planning). Conversely, these associations can be transformed in habits through extensive learning, where the conditional stimuli directly elicit responses without references to the outcomes to be obtained. More generally, this refers to the dichotomy (cf section 1.2.2) between

goal-driven behavior (where the behavior is driven by internal goals and can adopt complex schemes) and stimulus driven behavior (where the agent mainly reacts to perceived stimuli).

1.2.1 General loop of behavior

We have seen above that it is important to identify stimuli in the world as possible goals of behavior (emotional or pavlovian learning) and relate them to the corresponding need they can satisfy, to decide if it is worth triggering responses to get them (motivational or instrumental learning) [5]. Often the formulation of perception and action is considered as the beginning and end of a linear process, a stimulus - response arc. Organisms however have goals and agency, with behavior being more akin to a control loop, with inputs modifying outputs that in turn modify the next set of inputs to achieve an ecological life sustaining goal ("Those which are most useful to the organism"). Here, we refer to behavior as the internally coordinated response (external or internal actions) of living organisms to internal and external stimuli. With such consideration of behavior - as that which leads to responses monitoring a transition from one state to another, of the type Situation 1 - Action - Situation 2 (S1-A-S2), any behavior can be viewed as initially, and essentially goal driven. S1 can be interpreted as the initial condition eliciting A as the possible response and S2 as the consequence that can be anticipated if A is preactivated. Conversely, if S2 is the desired state, A is the response that has to be activated to reach S2, which is possible if S1 is compatible with the current state. Else, A can display a sustained activity, as in working memory (cf section 1.4.5), and remain actively waiting until S1 is satisfied.

More generally, throughout this thesis, we elaborate on behavior through these key ingredients :

Sensations : which can be internal or external; local or global (contextual), referred to as *Stimulus* for simplicity
The sensory cortex encodes this sensory information (called posterior cortex for simplicity, since most of them are posterior to the central sulcus

Response : which can be external or internal, in which case it is referred to as *Action* (thus this could be an internal action of updating the working memory)
The motor and premotor cortex encode information related to external responses, while the limbic frontal regions are responsible for internal responses (e.g. selection of goal and motivation)

Outcome : which relates to the consequences of actions, and refers to the reinforcing aspects of the pain / pleasure circuitry
These are encoded in the insula or medial temporal lobe, depending on the emotional and motivational aspects.

To illustrate this point, let us try to examine the ingredients of the basic behavior of hunger satisfaction (the brain circuits mostly involved in triggering this behavior are mentioned in parenthesis, as it will be more clear in Chapter 2), as different brain regions participating in the answer to one of four fundamental questions :

- *What* is the goal of my behavior ?
- *Why* should I spend energy satisfying the corresponding need (and upto which level) ?
- *Where* is this goal ?
- *How* should I behave (which response should I trigger) to get it ?

The perception or feeling of hunger (in the insula, hypothalamus) activates the motivation to eat (in the dorso medial prefrontal cortex (dmPFC; pgACC)). In order to trigger an appropriate action, first the progress towards the desired goal needs to be checked. If food is already present in the mouth, the animal needs to eat until satiety. This is checked by the ventro medial prefrontal cortex (vmPFC), which also represents the *value* of the expected reward. If the check is yes, then the sequence of actions for eating is triggered (masticating and swallowing in the motor cortex); if not, then it becomes a desired goal (in the lateral orbito frontal cortex (lOFC)). In the next sequence of actions, an intermediary goal is to get the food, and the cost to get the food is evaluated (in the vmPFC (sgACC)). The dmPFC (pgACC) compares the reward and cost to decide to act. In this case, food as a desired goal in the lOFC sets attention on all food like stimuli, sequentially evaluating the reward and cost until a choice is made. The cost, affordance or the goal itself is linked to comparing the strength of the goal, the urgency of the action and the confidence in the outcome.

In the example presented above, we see that even for the simplest of goals, a number of different regions of the prefrontal cortex (PFC) need to collaborate to achieve the said goal. We will focus on the role of each of the mentioned areas in the second chapter. However, all of these regions are not usually involved when we eat because more often than not, such behaviors occur under a normal, stable context, and this habitually used behavior is activated without much supervision. Nevertheless, the role of the prefrontal cortex goes beyond simply the ability to

navigate the sequence of reward and cost evaluation; its role is fundamentally to allow for the manipulation of existing behaviors.

On one hand, in simple and stable worlds, the most obvious solution becomes the dominant or default behavior, becoming strengthened after each success. Furthermore, after enough repetitions, certain actions are compiled together as chunks and directly trigger cached values, becoming stimulus driven. After a sufficiently long training period, they might become habitual, and rigid, where the current state is enough to directly trigger the response with no need to refer to a priori model of sensorimotor transitions or to the value of the outcome. On the other hand, in certain contexts, several close solutions to the same goal might be possible, or the dominant behavior might become erroneous due to a volatile and changing environment. This necessitates the need for cognitive control (cf section 1.3) in order to resolve the conflict between several possible responses, and the continuous need of an explicit deliberative system, that is able to track performance on multiple time scales, signalling the need to switch to an alternate or exploratory strategy. Said simply, to find the best global solution, contingencies between local decisions and their consequences must be evaluated, and corresponding reinforcements must be compared. This can also be associated to the domain of planning, with the classical steps of deciding for goal, motivation, strategy and execution, and of backtracking in the hierarchy when one step is impossible. We know that the brain has the capacity to assign control to the appropriate regions for each of these kinds of behaviors. In the following sections, we elaborate on the precise definitions of these behaviors, and review the current view in the field.

1.2.2 Stimulus driven vs Goal directed Behavior

Behavior has been traditionally divided into two categories : external, stimulus driven on the one hand, and internal, goal directed on the other hand. This traditional viewpoint induces two competing requirements - (i) to respond quickly to familiar situations while being able to (ii) adapt to novel ones and plan for the future.

Quickly responding to the immediate environment is accomplished when familiar stimuli activate well established neural pathways, producing stereotypical behaviors. These behaviors are executed quickly because they are based on stable, "concrete" rules (1.5.1), i.e. they are grounded on specific stimulus-response associations. Even though the acquisition of such associations is a slow and incremental learning process, the triggering of these associations is done with little need for control nor oversight. In this kind of behavior, stimuli (or states of the world) trigger actions, which then produce outcomes (new states of the world).

In contrast, sophisticated, goal directed behavior requires the ability to act on, or predict future events, rather than just react to the environment. Goals, or mental representations of desired future states of the world determine actions, lead to new states of the world, which can be compared with those envisioned as goals.

The aforementioned distinction does not constrain or formalize the neural or cognitive mechanisms that disassociate the two systems. In the field of machine learning, and more precisely, reinforcement learning (RL), this dual system theory of behavior has been formalized as a mapping to **model free** (MF) and **model based** (MB) forms of RL. While RL algorithms can be categorized along many dimensions, MB vs MF algorithms are contrasted based on the extent to which they represent the environment or not, and how they respond to changes in the environment or the agent's goals. MB algorithms maintain a representation of the problem beyond the state and action space, usually by including the transition and reward function. This internal model enables the agent to guide it's decisions by considering the consequences of its actions. These algorithms thus adapt more readily by leveraging the task model to dynamically plan toward an arbitrary goal. However, maintaining the task model and computing an action plan can quickly become intractable. On the other hand, MF algorithms do not maintain such an explicit model. Instead, they store a set of estimates, each representing the aggregated reward history of choices made by the agent in the past. While such algorithms cannot adapt as easily due to their strategy of integrating reward history into a single value estimate, they offer an efficient approach to learning and decision making [52].

It may be considered natural to identify the value of mapping model free RL to stimulus driven behavior since it is expressed in the stimulus response outcome form, and to map model based RL to goal directed behavior due to the internal model or cognitive style representation of the world. A MB strategy involves prospective cognition and assessment of the consequences of taking particular actions and supports the computation of value transformations when relevant conditions change - all features of goal directed behavior. Similarly, like stimulus driven behavior, MF systems are reactive, making future estimates only based on reward values encountered in the past and caching information about the utilities of those outcomes. However, such clear distinction is not always possible, and the two models might overlap. For example, standard MF algorithms augmented with additional computational machinery like a working memory, can mimic a MB planning strategy [138].

To the extent of finding neural correlates of these systems in the brain, it was the temporal difference (TD) learning mechanism, that RL is based on, which sparked a turning point in the understanding of dopamine function in the brain.

The phasic firing patterns of dopamine neurons in the VTA were shown to mirror the characteristics of a TD-RL reward prediction error. This has provided a bridge between behaviorally descriptive models and a functional understanding of a learning algorithm that may be supported by the brain. Thus, model based and model free (and similarly stimulus driven and goal directed) algortihms (in the context of learning and decision making), are better understood as the two ends of a high level process that emerges through the coordination of many separable sub-computations. However, these dichotomies may not be helpful in identifying those unique and separable mechanisms underlying behavior.

The aforementioned distinctions are useful in explaining behavior from the point of view of the observed input and response (SD vs GD), or as a computational formalism (MF vs MB). Nevertheless, they are insufficient to unravel computational and mechanistic primitives supporting behavior. In the next section, we review accounts of behavior in terms of *processing*. Shifting the focus to this axis may provide a more clear path towards finding computational primitives liking brain and behavior.

1.2.3 Automatic vs Controlled Processing, or Habitual vs Contextual Behavior

The view that human cognition may comprise two different types of processing, automatic and controlled, has been a theme in the psychology literature for a long time.

In earlier papers [184], an **automatic** process was defined as the activation of a sequence of nodes that nearly always becomes active in response to a particular input configuration, and that is activated automatically without the necessity of active control or attention by the subject. In general, automatic processes are thought to operate through a relatively permanent set of associative connections, requiring an appreciable amount of consistent training to fully develop. According to these authors, **controlled** processes are defined as those processes that rely on attention for execution, while automatic processes are those processes that can be carried out without attention [137]. For example, in the Stroop task (2.5.2) (where participants are cued to determine either the ink color (color naming) or the color name (word reading) associated with a word stimulus), color naming was considered to be controlled because it relies on attention. Without attention to color, subjects will read the word by default. Furthermore, the color has no impact on word reading, even when it conflicts with the word being read. Conversely, word reading is automatic because it does not appear to rely on attention. Even when asked to name the color, if a conflict word is present, it slows the response to the color.

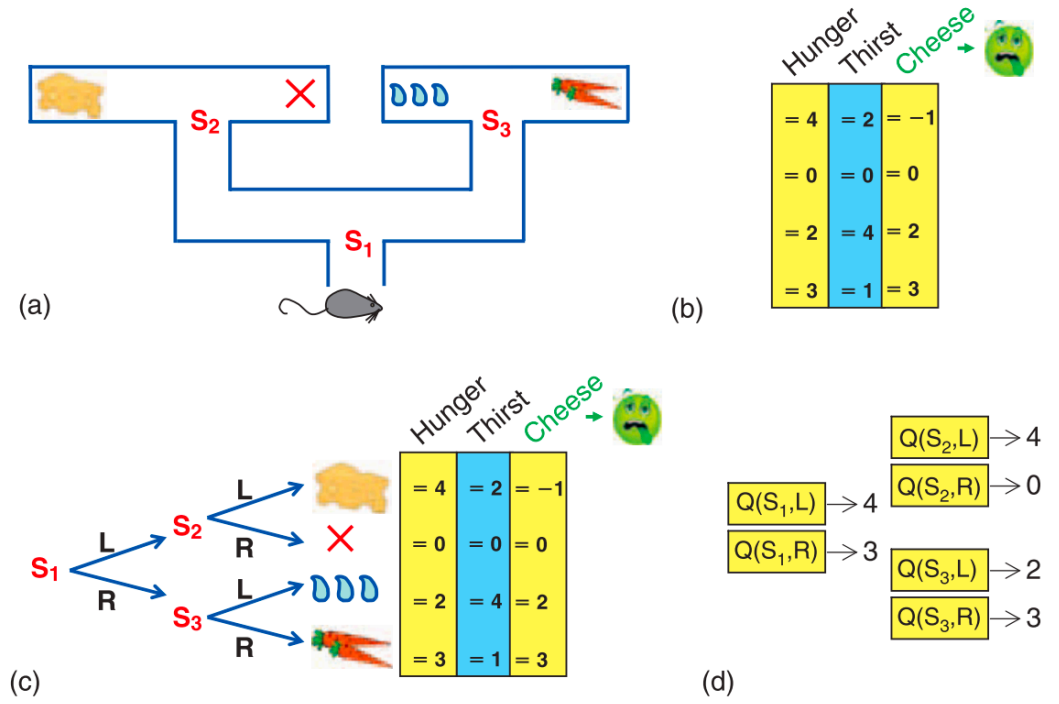


Figure 1.1: Model-based and model-free actions in a simplified maze task. **(a)** A simple maze with three states (S1, S2 and S3) from which the animal has to make left-right decisions, with the terminal states yielding outcomes of cheese, nothing, water or carrots. **(b)** The values of these outcomes under three different motivational states : hunger, thirst, and cheese devaluation. **(c)** A tree-based model of the state-action environment, which can be used to guide decisions at each state by a model-based controller. **(d)** The cached values available to a model-free, habitual controller. Immediately after cheese devaluation, these values do not change (in contrast to the model-based controller). It is only after direct experience with the devalued cheese that the value associated with Left (S2), and subsequently Left (S1), is reduced. Figure taken from [58]

However, there are problems with this simple dichotomous distinction. First, it is not clear that any cognitive process can occur entirely independent of attention. For example, although an individual is reading the words on this page, presumably they are not reading the words out loud. Second, it is not clear that color naming is always dependent on attention and control. In essence, such accounts propose that processes that demand control are distinguished from automatic processes, which involve associations sufficiently strong as to be resistant to distraction or interference [187].

All processes rely on attention to some degree, and this may vary in a graded fashion. Thus, some processes are more automatic than others, and processes vary in their automaticity based on the context in which they occur. This axis of automaticity starts from coordinating conflicting responses demanding contextual control on one end, to readily triggered default behaviors that can turn into rigid unchangeable habits on the other end.

Default behavior : Routine sensory conditions tend to elicit some kind of default response, which might either arise due to some kind of innate knowledge about the world, resulting from an evolutionary process (think about the reflex action of flee or freeze when a prey encounters a predator), or due to prior learning which has been reinforced repeatedly.

In the initial stages of habit learning, behaviors are not automatic. They are goal directed, such as for example, an animal working to obtain a food reward. When goal directed behavior consistently produces the same behavior in response to the same stimulus, that behavior is likely to become a habitual response. Once a habit is formed, behavior can become inflexible in the face of changes to the environment that render it no longer desirable. In experimental settings, with extended training or training with interval schedules of reward, animals typically come to perform the behaviors repeatedly, on cue, even when the value of the reward to be received is reduced so that it is no longer rewarding (for example, if the animal is tested when it is sated or if its food reward has been repeatedly paired with an aversive outcome). Habits are promoted in situations where the contingencies between action and outcomes are weak. Particularly, when the values of actions are divorced from outcomes, habitual control takes over. Dickinson defined the goal oriented, purposeful, non habitual behavior as action outcome (A-O) behaviors and labeled the habitual behaviors occurring despite reward devaluation as stimulus response behaviors (S-R). From simple motor actions to choice of meals, travels and exercise routines, behaviors become more automatic (faster, more accurate, less susceptible to interferences) the more often those behaviors are performed in the presence of a particular set of cues.

In the consciousness literature, the study of habits invokes this dichotomy between the conscious, voluntary control over behavior, which is considered the essence of higher order deliberative behavioral control on one hand, and lower order behavioral control that is scarcely available to consciousness, on the other hand. To consider the defining characteristics of habits, through this lens, four key points become evident [86]. First, habits are largely learned i.e. they are acquired via experiment dependent plasticity. Second, habitual behaviors occur repeatedly over the course of days or years, and can become rigid and fixed. Third, fully acquired habits are performed almost automatically, allowing attention to be focused elsewhere. Fourth, they tend to invoke an ordered, structured action sequence that is prone to being elicited by a particular context or stimulus. Thus, these characteristics suggest that habits are sequential, repetitive, motor or cognitive behaviors elicited by external or internal triggers that once released, can go to completion without constant conscious oversight. In addition, habits can be defined experimentally as being performed not in relation to a current or future goal but rather in relation to a previous goal and the antecedent behavior that most successfully led to achieving that goal.

Coming to the experimental literature, this distinction between action-outcome vs stimulus-response systems is an ongoing debate [71] [163]. Evidence suggests that these are not independent "systems". For example, after training that produces habitual behavior in rats, goal oriented behavior can be reinstated if the infralimbic prefrontal cortex is inactivated [54]. This finding has led researchers to believe that the circuits controlling goal oriented behavior may be actively suppressed when behavior becomes habitual. The idea that there is a dynamic balance between control systems governing flexible cognitive control and more nearly automatic control of behavior responses supports the long standing view from clinical studies that frontal cortical inhibitory zones can suppress lower order behaviors. This view has become especially important in models of system level interactions [57].

The distinction described above and reiterated in scientific literature, between controlled and automatic processes is useful in shedding light on two important mechanisms that are involved in cognition - (i) attention to relevant features (or on the contrary, ignoring irrelevant ones), and (ii) strength of action - outcome associations based on context . Yet, it still does not provide a clear picture of how basic learned behaviors might be manipulated and coordinated.

The study of learning and decision making, in both human and nonhuman animals has broadly been studied under the idea that behavior is governed by two separable controllers. Behavior has thus been dichotomized against several dimensions, including emotion (Hot/Cold), action selection (habitual/goal-directed), judg-

ments (associative/rule-based) and reinforcement learning (model-free/model-based) [52]. Critics have pointed to the multitude of dual processing accounts, the vagueness of their definition, and the lack of coherence and consistency in the proposed cluster of attributes for the dual system accounts. Largely, these dichotomies are subsumed under the terms System 1 and System 2 thinking. In the above sections we have tried to disentangle the defining versus the correlated features in some of these accounts. In the following section, we focus only on the control aspect of behavior - how it might arise, its minimal defining characteristics, and the neural underpinnings that support the implementation of such control.

1.3 Cognitive Control

To grasp an intuitive understanding of how Cognitive Control is implicated in behavior, let us look at a few examples : two based on daily life, and two based on laboratory tasks. Consider the sequence of actions required for driving. Any experienced driver understands and follows standard rules without much thought or effort - when to stop at traffic signals, when to accelerate and break and what side of the road to overtake from. Until this driver visits the UK on holiday, and now needs to modify some of these rules for driving because the context (environment) has changed - the driver needs to change the side of the road (s)he drives on, while still following most of the other basic rules for driving (the right sensorimotor actions for changing gears, etc). The driver also understands, or in able to predict the consequences of not adapting the driving rule : one might need to pay a hefty fine, risk getting into an accident, or even be questioned by the police, and hence there is a significant cost attached to the failure to adapt. Alternatively, consider our instinctual response of looking for our phones and swiping towards the green button when we hear a phone ring (which may or may not belong to us); until we find ourselves in a meeting and this behavior needs to be suppressed. Paying attention to a phone ring is in fact a habitual behavior, even if answering it might be the dominant behavior. In either case, it requires an active suppression of the habitual tendency, which we know to be maladaptive in certain social contexts. In controlled, experimental conditions, we can consider the standard Wisconsin card sorting test (WCST; cf section 2.5.4), where one has to go on following an underlying rule of matching the presented card to a number, color or shape, until the experimenter changes the task criteria. The participant will subsequently start to make mistakes and will need to reconsider the rule to be applied. This requires not only inferring a rule that is not clearly specified (as a latent state), but also an active monitoring of one's performance, such as to be able to detect when there is

sudden drop and must be attributed to external factors. Yet another example is the Go - NoGo task in which the subject has to mostly make the same response unless the stimulus presented has a certain property.

In each of these cases, there seems to be a default or dominant way of doing things and performing actions, until something in the environment changes and the underlying default (or usual) rule needs to be adapted or suppressed in order to cope with the change in the environment. The adaptation requires one to maintain the new context in mind (working memory), selectively attend to a few features in the environment (selective attention) and monitor one's belief about the current state in the world (performance monitoring). This ability of inhibition or suppression of a default response, in favor of an alternative behavior more suited to the context and environmental changes is generally termed as "cognitive control". It is an ability that both humans and non humans demonstrate ubiquitously, and perform with much ease, all the time. This kind of flexibility, at the most basic level requires coordination between motor control and a higher level cognitive control. Here, we refer to motor control as the ability to plan and select specific actions while cognitive control is the ability to coordinate a set of responses. At the neuroanatomical level, it can be thought of as the different roles of the agranular and granular frontal cortex [165], with the agranular cortex implementing the default behaviors and responses and the granular acting on top to selectively inhibit these responses.

In summary, goal directed behavior requires the selection of task relevant information and the suppression of task irrelevant noise. A prominent characteristic (as seen from the examples above) of most current theoretical models of cognitive control is that this process is mediated by higher level control signals that bias the state of lower level neural processing. The origin of these control signals is commonly seen as originating in the prefrontal cortex. In the context of action selection, this top down control is particularly needed during situations of response conflict, where a predominant response (default behavior) needs to be inhibited in favor of an alternative response or no response at all. This corresponds to the view of cognitive control that the rest of this thesis is based upon : a set of cognitive processes which act not on specific stimulus response contingencies but on the performances of previously learned, more basic behavioral modules (what we call Rules, cf section 1.5).

A core problem here is what is referred to in literature, as the stability-plasticity dilemma, which is the degree to which a new instance (rule) should alter existing knowledge about a class of instances (schema, or here, Task Set) without destabilizing such knowledge. The cognitive tasks we perform at each moment, and the efficacy with which we perform them is a result of a complex interplay of deliberate

intentions that are governed by internal goals (endogenous control) and the availability, frequency and recency of the alternative tasks afforded by the environment and its context (exogenous influences). Effective cognition requires a delicate, 'just enough' calibration of endogenous control that is sufficient to protect an ongoing task from disruption but does not compromise the flexibility that allows the execution of other tasks when appropriate.

Some authors have defined cognitive control as the general capacity to use an internal contextual representation to guide full pathways of thought and action, in accord with goals [13] [112]. In particular, it has been described as the ability to perform task relevant processing in the face of distractions or in the absence of environmental support, specifically by active maintenance and flexible updating of task representations over time, in order to pursue task relevant objectives and goals [30] [70].

Elsewhere, this ability or capacity for control is interpreted as a set of superordinate functions which include working, semantic and episodic memory; perceptual attention; and action selection and inhibition [24]. The state of control at any instant can be characterized in terms of *(a)* its direction, the specific task objectives toward which control is directing subordinate systems; and *(b)* its intensity, the strength of its top-down input to those systems.

It is evident that allocation of control depends on the circumstances of cognitive demand (which can be interpreted as goals, context, or task depending on the granularity of the definition) and as such requires the dynamic recruitment of available cognitive processes that can appropriately meet these demands. In this view, cognitive control can be though of as having the following features :

Regulation : The capacity to govern or influence lower level information processing mechanisms

Specification : A decision on which controlled tasks should be undertaken and on how intensively they should be pursued

Monitoring : The system must have access to information about current circumstances and how well it is serving task demands. It refers to the three criteria to monitor once a goal has been selected : the outcome to reach, the action to perform and a measure of progress

Having established the defining features of cognitive control, and the general set of functions it uses, there are three focal questions that a computational modeling study of cognitive control might address [24] :

- How do control functions influence information processing ? What are the top down regulative effects of control ?

- How might control emerge from learning and experience ? What are the bottom up factors that govern the selection and modulation of top down control signals ?

- How does the architecture and representation of the control system come to assume its specific form or functional organization ?

This thesis is an attempt to answer each of these three questions systematically, through biologically plausible computational modelling and coming to a consensus by reviewing the extensive neuroscientific literature on cognitive control and its associated mechanisms and constructs. Figure 1.2 explains this gradient of control we have described thus far, schematically. Specifically, the full capacity of control that this figure describes requires incrementally such constructs as Goal, Context, Plan, and monitoring. Some of these of have been elaborated in previous sections. In the next section, we begin by delving a little deeper into the constructs and mechanisms that support effective complex, hierarchical control (as seen in primates) i.e. working memory, attention and performance monitoring.

To talk about cognitive control, we need to first talk about *what* this control is exerted over. In any scenario, there are possibly infinite kind of actions that can be taken, any number of different behaviors that may be demonstrated. Only some, if not of those is ultimately chosen, while the others are suppressed. Since we carry within us a repertoire of possible behaviors, it is reasonable to start with the smallest currency of any well defined behavior, by decomposing it into simple routines for selecting actions (or thoughts) - a rule.

After introducing conceptually, the structures and mechanisms needed to understand cognitive control, in the next section, we try to narrow down the definition of a "rule" and its related aspects and categorizations (concrete vs abstract, implicit vs explicit), from the view point of how these constructs are implicated in cognitive control. In the second part of that section we discuss the conceptualisation of abstraction, and theories of how the brain might be organized to support such abstraction with hierarchical organization. We also look at behavioral evidence of rules in animals, and how the definition of a rule in cognitive science is different from that in computer science and logic.

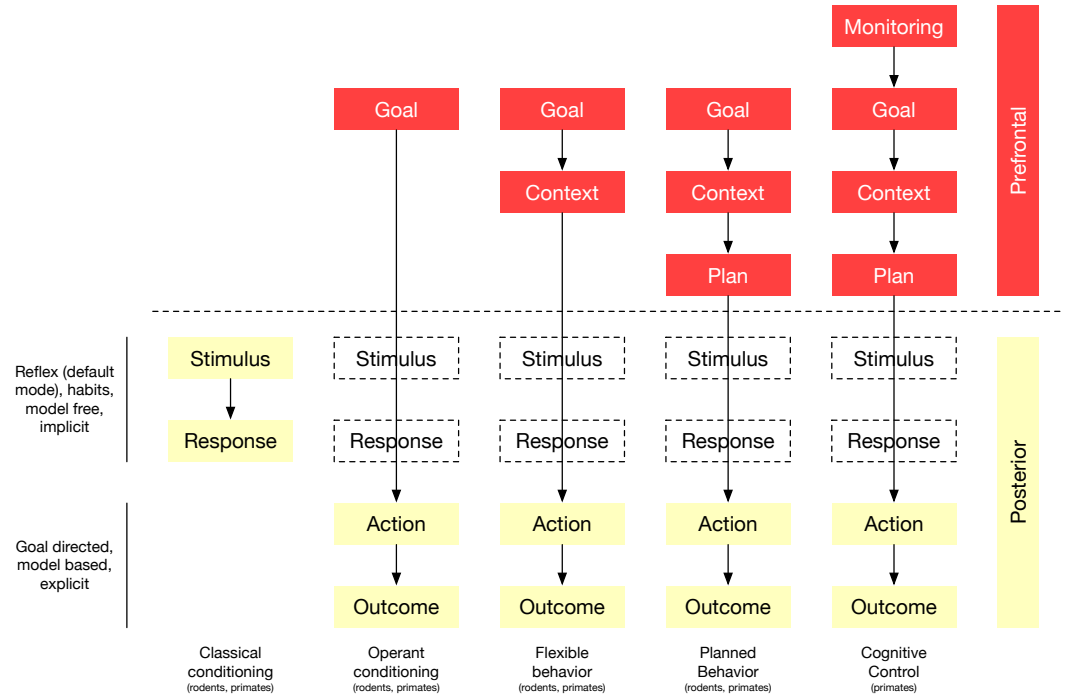


Figure 1.2: **Schematic figure illustrating the gradient of cognitive control** Starting from Stimulus-Response associations acquired through pavlovian conditioning which can be seen as unchanging behaviors (habits or default behaviors) that need to be overcome by exerting control. Here, a distinction is made between Responses (external actions which are triggered by the external environment), and Actions (voluntary responses that can be internally or externally directed). Next, are Action-Outcome associations learned through instrumental learning paradigms, which are sensitive to reward outcomes, generally controlled by certain goals that are intrinsically rewarding to the organism. Slightly more flexible behaviors require the organism to make contextual representations (external as in environmental cues or internal latent states) of the learned associations such that a change in the context is able to trigger a change in the expression of an underlying rule (contextual control leaning). Planned behaviors further involve the organization of sequential behavior in time, that can be triggered by goals (or sub goals) such that once initiated, will reach completion without explicit oversight. Finally, a complete and complex manifestation of cognitive control requires all the above ingredients, in addition to a contextual monitoring of motivations, goals, errors, and self performance so as to recognize uncertainties in the environment and link them to the correct causes.

1.4 Mechanisms and structures involved

1.4.1 Constructs

Studying cognitive control from a computational frame of reference, calls for oft used terms in the literature. Below, we briefly describe such constructs, and emphasize the usefulness of such concepts.

Goal

Goals in general refer to desired or anticipated outcomes or end states. These goals can be the consequence of physiological needs such as thirst and hunger, as well as various other needs or motives such as making friends, acquiring knowledge, etc. Goals are critical because the decision process is assumed to be intended, and perhaps even optimized, to achieve them. Indeed, optimality must be assessed only in the ecological context of a goal. Thus, behavior that is "suboptimal" with respect to certain objective goals such as maximizing accuracy might in fact be optimal with respect to the idiosyncratic goal(s) of the decision maker.

State

At the root of computational models (especially RL) is the concept of state representations, an abstract representation of the task that describes its underlying structure. States can be tied to external stimuli, or they can include internal information that is not available in the environment and must be retained in memory or inferred, such as one's previous actions, or the context of the task.

Value

Humans and animals make predictions about the rewards they expect to receive in different situations known as value representations. They drive choice : expected value of available options are compared to one another and the best is selected. They support learning : expected values are compared to rewards actually received, and future expectations are updated accordingly. This assumes the notion of common currency to be able to compare things.
Value is the subjective costs and benefits that can be attributed to each of the potential outcomes (and associated courses of action) of a decision process. Value can be manipulated by giving explicit feedback or monetary rewards to human subjects or preferred food or drink to nonhuman subjects. Value can also reflect more implicit factors such as the costs associated with wasted time, effort, and resources.

Context

Contexts encompass internal (cognitive or afffective), interoceptive and external settings. Thus context frames and contains a priori information about reward. Consider the example of the flat vs hierarchical rule discussed in [14]. When the rule is flat, all the cues that play a role in the decision play the same role and are combined at the same level. With a hierarchical rule, we make a difference between the lower cues which are the main cues and the higher cues which are put at a secondary level and can be considered as context. Context is not a property per se, but depends on the situation; of course, of this 'secondary' cue corresponds to the background which is effectively often secondary in the decision process. In this case, it is also possible to propose the fact that context is more to define the kind of task rather than the main cue that will trigger the action.

1.4.2 Uncertainty : Stochasticity vs Volatility

Uncertainty typically arises in a situation that has limited or incalculable information about the predicted outcomes of behavior. Uncertainty can be induced not only by lowering the probability of stimulus-response-outcome contingencies, but also by fundamental changes in these contingencies that forces a modification of previous beliefs. It can arise from :
Stochasticity or expected uncertainty [7]; it is inherent in the decision making process; S-R-O rules learned from past events are weak predictors of the outcomes of future actions and this unreliability is known and stable.
Volatility or unexpected uncertainty; it arises from fundamental changes in the S-R-O contingencies of the environment that invalidate predictions based on previous experience; it is a variation in the frequency of the changes in existing S-R-O contingencies across time

1.4.3 Task Set

In daily life, humans are constantly confronted with a lot of information from the environment, but only part of this information actually gains access to the cognitive system. Furthermore, we also have a very large behavioral repertoire of likely responses we can make, given even this partial information. Consider the standard example of standing in your kitchen, prepped to prepare a meal. First, you inadvertently only a few objects of focus - one of them being the vegetables on the kitchen counter. Now given this filtered information from the environment, your likely response is to chop the vegetables. On the other hand, in the context of a grocery store, the object of information from the environment - the "vegetables" -

your response is very different, ie, it is to put them in your grocery basket and head to the checkout counter. It is evident that depending on the context (and in the example described, the context is the task), we pre-select all the behaviors (or rules) that would be pertinent here, and then user finer processing to select one among them. Furthermore, at least most of the time, it is relevant information that is attended to, whereas irrelevant information can successfully be ignored. Two questions thus arise : first, what decides which information is relevant and thus gains access to further processing, or which information is to be discarded as irrelevant. And second, what structures might support the sub-selection of valid behaviors from the repertoire of all behaviors in long term memory ?

One possible answer to these questions is the conception of "task sets" or task representations. Such representations can form the basis of modulating which stimulus information is processed and which is not, by narrowing down the focus of attention towards the relevant set of behavorial responses. Thus, a task set can be thought of as a set of all the stimulus - action - outcome associations that are pertinent for the task at hand.

In the experimental literature, to perform a specific task, humans are thought to enter a task dependent cognitive state, mode, or set that is maintained for the duration of the task [66]. A task-set (TS) is a configuration of cognitive processes that is actively maintained for subsequent task performance. It is the representation of a mental state corresponding to any currently used rule-mapping in order to perform a given task [180]. When subjects are given instructions prior to a task, after practice for several trials, the task information is maintained as a configuration of perceptual, attentional, mnemonic, and motor processes necessary to perform the task. When subjects are then asked to perform another task, they have to establish a new task set in a form distinct from the previous one. The stimulus set and response set may be same between different tasks, but the *rules* of association between the S-R might differ. A TS has to be specific in the sense that it represents a rule of a specific task to be performed. It is nonspecific in the sense that it can be applied to any stimulus as long as it belongs to a task-relevant stimulus set.
Action sets - the different ways to do the same thing (eg cutting the vegetable with this or that tool, with the right or left hand etc)

The prefrontal cortex is associated with a variety of executive functions, including this ability to flexibly change cognitive configurations (ie task sets) to newly relevant task demands. Theoretically, there may be two different aspects to the process of establishing a task set. First, the relevant rules must be activated (eg : [136], [174]). Secondly, interference from competing task sets has to be minimized, possibly through active inhibition [136]. The dissociation of these two functions

however is fairly complicated [104], and we talk about it in more detail in later sections.

The updating of a task set may be a self-regulatory mechanism emerging from interactions and competitions among rule representations. The key feature of a task set is its prospective and predictive nature. In this sense, the task set is a crucial concept in elucidating the causal mechanisms of the brain in creating complex behaviors and abstract thoughts. Task-sets not only help in avoiding mistakes in behavior, the benefit of task set is facilitation of subsequent task performance. Many task-sets which are initially acquired through instruction or trial and error, are stored in our memories. The more we practice a task, or the more recently we have practiced it, the easier it becomes to reenable that task set. To investigate the neural correlates of a task set, experimenters typically record brain activity while subjects are required to perform a specific task; the idea being that such task set activity likely determines the activity in areas involved in task execution. Single unit studies have shown that representing, updating, and implementing task sets are subserved by interactions among different sets of neurons in the same region of the prefrontal cortex. By contrast, imaging studies have shown that each of these processes is subserved by distinct regions in the prefrontal cortex and other areas. We discuss these findings in greater detail in Chapter 2, but these results suggest parallel processing at different levels of brain organization ie at the neural level, the implementation of task rules, task items and responses are distributed.

In the following section, we elucidate on three core mechanisms that enable cognitive control : attention, working memory, and the monitoring of errors.

1.4.4 Attention

Real world choices are typically guided by multiple shifts in attention between the choice and alternatives. Without the ability to narrow down focus to a subset of environmental features (as input) and of behavioral responses (as actions), conscious awareness would be completely swamped by irrelevant information, or one would never be able to resist inappropriate urges. Given that cognitive control is primarily the ability to inhibit behavioral sets, attentional control is the ability to filter information that enters the decision circuit.

Attention might be the brain's solution to the problem of living in an informationally dense world with a limited capacity for processing information [199]. Only some environmental information, thoughts, and sensory inputs can be processed at any point in time and attention is the mechanism by which processing priority is accomplished. Not all sensory input is behaviorally relevant and processing irrelevant information can be energetically very costly. Attentional selection is

the process of selecting information for prioritized processing in order to demonstrate adaptive behavior that is improved upon as we learn information about our environment that relates to our goals, and thus learning and attention are tightly linked.

Selective attention can be directed by two types of mechanisms : voluntary (top-down) and involuntary (bottom up), which is the distinction between goal directed (top down) and salience directed (bottom up) mechanisms. In the absence of explicit instructions, which is normal experience in everyday environments, the control of selective attention needs to rely on internal mechanisms that dynamically track the relevance of sensory information in the environment.

According to the representation account, attention is the modulatory influence that representations of one type have on selecting which (or to what degree) representations of other types (for example, the rules to be triggered or inhibited) are processed, that is, how representations on one type (for example, the representation of active goals) guide the flow of activity among other types [137]. An attentional set (like a task set or action set) is a definition of the representations of the advanced information involved in selecting task relevant stimuli and responses. By determining which information enters the decision circuit, attention affects the temporal dynamics of several decision related computations, including stimulus identification, valuation, comparison to previously attended alternatives and action selection.

One view of attention is that it consists of separable, yet interconnected brain networks that influence computational priority, controlling what information enters conscious awareness. According to another theory, the objective of cognitive control is the prioritization of computations of specific input information so that uncertainty can be minimized, and attentional functions operate to serve cognitive control in the reduction of uncertainty in temporal, spatial and process/response domains. Alerting increases the predictability in time of the upcoming information that is to be processed. Orienting acts to select the most relevant and important information, in space, to be processed. Orienting and executive control of attention are differentiated in that orienting acts at the input stage to filter (or attenuate) task irrelevant information whereas executive control acts to bias the task relevant process at the processing and response stages when there is competition between processes.

1.4.5 Working memory

Working memory, on the other hand, is the *sustained attention* or *active representation* of a limited amount of currently relevant information so that it is available

for use; the key word here being sustained. This sustained activity can be used to maintain attention for a long time, solving the problem of holding a 'goal, task or context' in mind till the time the lower order processing required for action is finished. The term 'working memory' has traditionally been associated with the maintenance of specific stimulus features, such as phonological, visual or spatial information. However, the concept of working memory and the mechanisms that enable active maintenance of stimulus features can be extended to other, more abstract types of information. For example, sustained activation of an individual's current environmental context and task goals is also necessary for flexible behavior and higher cognitive processing.

Originally coined by Newell and Simon (1956) in the context of computer science, the term working memory was introduced into cognitive pyschology by G. A. Miller, Galanter, and Pribram (1960) who used it for the idea of holding goals and subgoals in mind in the service of planning and executing complex behaviors (Cowarn, 2017) [92]. A central idea behind most neurobiologically based computational models of WM is that neural activity can be sustained through mutual excitation, where populations of interconnected neurons send each other excitatory activity in a self sustained manner (recurrent activity). Computationally, this corresponds to a stable attractor in a dynamical system : a state that remains constant over time once the system enters the vicinity of that state. This mechanism of working memory is specifically described as robust active maintenance. Functionally, the ability to robustly maintain activity over time must also be complemented by an ability to rapidly update to encode new information into working memory. These two demands are mutually contradictory, and the concept of gating has been introduced as a way to dynamically switch between robust maintenance, versus rapid updating [193]. There are other proposed mechanisms, like the possibility for certain neurons to keep a long plateau activity thanks to internal mechanisms. Sustained activity can be seen not only in the PFC but also in higher perceptual cortical regions and in other neuronal structures like the cerebellum.

Working memory dynamics, and models thereof, include how sequences of events are temporarily stored in ventrolateral and dorsolateral prefrontal cortex, how these sequences are unitised, or chunked, into cognitive plans and how interactions of prefrontal regions with other brain regions enables predictions and actions to be chosen that are most likely to succeed based on sequences of previously rewarded experiences.

1.4.6 Prediction of errors / Monitoring of errors

Recent years have seen the emergence of a wave of influential theories that highlight the predictive nature of cognition. A common denominator of these theories is that they paint a picture of the mind wherein our mental representations of the world become active before we engage with reality (top down processing); this view contrasts with traditional perspectives that assumed that our representation of the current state of the world emerged only after we have acquired evidence from our sense organs (bottom up processing) [84].

A prominent theory within this framework is the Predictive Processing approach that argues that every encounter we have with reality is akin to scientific hypothesis testing. For example, a person who is about to open the fridge already has prior representations of what they are about to see (a milk carton); to the extent that this representation successfully predicted the event to come, there is no need for much additional cognitive processing; however, when a discrepancy between the prior representation and bottom up inputs is detected (if there is no milk left, a prediction error), then there is a need to update the mental representation in light of the new evidence.

Missing from the picture until now, however, is the process by which organisms detect environmental change and begin the process of either switching between or learning new behavioral strategies. Such a process should successfully handle changes in both the statistical parameters of the environment (alterations in volatility, outcome probability and outliers) and its contingency structure (changes in the state space, its transition properties, and the introduction of new events).

Conflict between Task Sets : Task switching is the process of selecting between competing task sets. Because a task set is a configuration of perceptual, attentional, mnemonic, and motor processes, the conflict can occur at various stages of cognitive processes depending on the differences between taks, which could be associated with activation in different brain areas. Compared with conflicts at stimulus and response processing stages, conflicts at a conceptual rule-processing stage occur regardless of the overlap among stimuli and among responses.

1.5 What are Rules?

"Rule" is a commonly used word in everyday language : ranging from rules of a game to imperceptible social rules like being polite to acquaintances. Given the intuitiveness of its use, in this review, we hope to provide an operational definition of "Rule" in the context of understanding the neuro-computational properties that lead to behavior (animal or human), starting from the most generic and proceed-

ing to the most specific. We will borrow and use literature from logic, psychology and neuroscience to get a rounded view of what rules are, how they might be implemented and processed in the brain and how we adaptively use them in our lives.

A note on nomenclature : There are laws (or rules) of physics (eg, gravity ie an object released from a height will fall to the ground) and social and moral laws (eg, being polite is requested to be successful in certain circumstances). Such rules exist outside of us and our influence. We don't invent them and yet our behavior conforms to such rules and they play an important role when we define our (internal) rules that guide our behavior ie behavior that we have some agency over. More precisely, we take into account the rules of physics when we define our own sensorimotor behavioral rules (in the motor and premotor cortex) and the natural and social laws when we define our emotional or motivational behavioral rules (in the limbic prefrontal cortex). The first kind of rules are factual or objective, i.e, they can be categorized as right or wrong while the latter rules are subjective, i.e., they can only be categorized as good or bad. On the other hand, more abstract rules such as the rule of addition, and other mathematical axioms, or syntactical rules of language are also beyond the scope of this study. Instead, we focus on simpler studies and tasks (despite the complexity of human behavior) simply because much of what we know about the encoding of rules in the brain comes from animal studies, which in itself require several constraints.

The Merriam Webster dictionary describes Rule[s] as prescribed guide[s] for action. In logic, rules can be transformations, as in , or simply expressions that specify a particular set of relations. For example, in propositional logic, *modus ponens* is a rule of inference that states "If P, then Q. P is true, therefore Q must also be true." To put it more generally, a rule then has two parts : a condition that must be satisfied to trigger the rule, and an action that follows once the rule has been triggered. The same holds true in the real world, and depending on the specifications of the triggering condition, some situations will trigger the rule in a clear prototypical fashion while others will partially match the conditions and will result in a slow and uncertain application of the rule.

In modern psychology, rules derived from formal logic are combined according to a reasoning program for using the schemas; a basic universal routine and a set of acquired strategies to account for individual differences. In the domain of neuroscience, we can think of rules as instances of repetition of spatio-temporal rela-

tionships between discrete object features, events and actions present as statistical regularities in the environment. Through repeated experience of these regularities, animals and humans learn to generalize and link these events and objects together as associations. Consequently, a triggering condition or situation can then extract and apply these rules and responses.

In a broad, general sense, rules are those beliefs, strategies, stimulus-response behaviors, influenced by memory and learning, reward-expectation based probabilities that guide decision making on the basis of prior learning recalled, via perception of current environmental states as a guide to action. First you store rules for sensori-motor (in the general sense, on the premotor and limbic sides) behavior. Then this memory provides a representation that can be used in anticipation (strategies) or be associated to their level of confidence (beliefs).

To make things more precise, in this section, we try to break down these different components of what constitutes the formulation of such "rules".

1.5.1 Concrete vs Abstract Rules

In the simplest case, when an object's properties, usually through a discrimination of one or more features of the stimulus, directly indicate a response or category, such associations are referred to as **concrete** rules [75]. They describe simple spatio-temporal links between objects, events and actions. These links are known as stimulus-response or stimulus-outcome associations [133]. They can discriminate on a single feature or a subset of features shared across a group of sitmuli. They can also discriminate using multiple features of a stimulus. For example, when I encounter a red traffic light while driving, I should press the brake. In this case, there are two relevant stimulus features (traffic signal, red light) and one state based condition (while driving). The formation of concrete rules is experience dependent, meaning that they are typically learnt gradually across multiple instances of experienced positive and/or negative reinforcement.

Abstract rules are rules that do not assign specific responses to stimuli but instead are used to select a set of concrete rules from all possible set of concrete rules that could be applied to the stimulus. For example, while sorting coins, concrete rules are the individual rules of sorting by size or by value, while the contextual decision of using the 'sorting by size' rule is an abstract rule. Abstract rules are complex and applicable to multiple exemplars. They describe interactive and causal associations between objects, events and responses. They require linking together different concrete rules and integrating them with instructions and information for achieving a particular goal, and under a certain context [133].

1.5.2 Abstraction of Rules

Humans and nonhuman animals show the ability to learn about and act on the perceptual relations between events, properties and objects in the world, but humans go a step beyond since they are able to grasp the higher order relation between these perceptual relations in a structurally, systematic and inferentially productive fashion. In that context, analogical reasoning is a fundamental and ubiquitous aspect of human thought. It is at the core of creative problem solving, scientific heuristics, causal reasoning and poetic metaphor. Humans form general categories based on structural rather than perceptual criteria, find analogies between perceptually disparate relations, draw inferences based on the hierarchical or logical relation between relations, cognize the abstract functional role played by constituents in a relation as distinct from the constituents' perceptual characteristics, or postulate relations involving un-observable causes such as mental states and hypothetical physical forces [156].

The understanding of abstraction from both philosophical and cognitive theoretical perspectives has evolved over time. In his Essay Concerning Human Understanding, John Locke distinguishes between particular and general ideas. Particular ideas are constrained to specific contexts in space and time. General ideas are free from such restraints and thus can be applied to many different situations. In Locke's view, abstraction is the process in which "ideas taken from particular beings become general representatives of all of the same kind" by dint of the mind's removing particular circumstances from an idea [122]. A theory advanced by Barsalou [16] advocates for a theory wherein connections between concrete and abstract concepts are direct and nonmetaphorical. In this theory, concepts take the form of simulators, which are semantic clusters that can generate infinite further examples of a concept. As we encounter examples of objects, we encode their perceptual features and store them in our memories. These features will form into clusters, which eventually become simulators, with a frame of previously encountered common features and a set of infinite possible simulations that the frame can generate. For instance, our perceptual experience with various chairs has helped us form a concept of '"chair". We can now use this concept to simulate infinite further examples of chairs.

In more recent studies, neuroscientific evidences of abstraction has been found. Several studies on nonhuman primates have shown that prefrontal cortex plays a

key role in abstract rule-guided behaviors [203], [146], [34]

By defining abstraction as the ability to move away, in a graded fashion from perceptual stimuli to mental constructs, it is useful to look at a few examples where several different types of "abstraction" can be seen . Contexts can generalize over more rules (**policy abstraction**), more dimensions may need to be integrated to make a decision (**relational integration**) or contexts may need to be sustained over longer periods of time while lower order decisions are made (**temporal abstraction**) [14] :

- Contexts can generalize over more rules, called **policy abstraction**
  If you are asked to raise your left hand when a red stimulus appears and to raise your right hand when a green stimulus appears, the rule is the specific association between the stimulus and the response. Alternately, the rule can be more abstract. For example, if you are asked to press the left button when two pictures are the same and to press the right button when the two pictures are different, the rule is not associated with any particular feature of the sensory stimuli. This "same or different" kind of abstract rule is often studied in primate experiments.

- We may be required to integrate multiple dimensions to make a decision - **relational integration**
  Consider two sequences of words of objects : *little-medium-big* and *light-shaded-dark*. A first order relation in this case might be represented as *bigger*(x,y) and *darker*(a,b). A second order relation linking these two together could be represented as *greater*(size(x), size(y)) and *greater*(shading(a), shading(b)).

- In most everyday tasks, contexts (or goals) may need to be sustained over long periods of time while lower order decisions are made - this kind of abstraction can be called **temporal abstraction**
  There are tasks that require an abstract hierarchy of goals, which need to be processed either sequentially (in the case of making coffee, with sub-goals being to get a cup, add sugar etc..), or need processing resources between concurrent tasks such as listening and reading, and remembering and returning to where you left off reading

Coming back to the hierarchical representation evoked above, contexts are more easy to define : At the first level of the hierarchy, corresponding to concrete rules,

the conditions correspond to sensory cues. At the other higher levels, they correspond to contexts (as manipulated in abstract rules for cognitive control). The types of abstraction mentioned above, can also be understood in terms of building in the PFC, different kinds of abstract rules, ie the what, where and when, mentioned above. Since contexts, in this definition, correspond to the 'sensory cues' used for abstract rules in the PFC for cognitive control, we propose that contexts could also correspond to the context of tasks, as they are mapped in this PFC, and particularly in the vmPFC and vlPFC when we are looking for the semantic representation of this context.

One theory by [83] suggests that there is a learning continuum in which a gradual abstraction occurs. On one end, there are initial conservative, fully concrete mappings (context specific representations), which paves the way for a kind of analogical mapping in which a relational structure is imported to a new domain with no support from object matches to a fully abstract mapping in which base domain contains variables and the target contains objects. Other researchers have argued that human subjects possess a qualitatively distinct system for reinterpreting sameness and difference in a logical and abstract fashion that generalizes beyond any particular source for stimulus control [156].

If two objects are not distinguishable in any sense by the observer, there is only a single object in mind, hence there is no abstraction. The requirement of distinguishability means that abstraction involves having at least two dimensions in mind : one dimension on which the stimuli differ, and another dimension on which they will be considered identical. Thus, when performing an act of abstraction, one makes a decision on which dimension is central, and by doing so, one designates other dimensions as secondary or irrelevant in the current context. Because abstraction entails selecting / attending to one dimension and disregarding other dimensions that might be salient, many acts of abstraction likely rely on cognitive operations often referred to as cognitive control and selective attention.

Aspects of abstraction

Compositionality Compositionality is the idea that new representations can be constructed through the combination of primitive elements [121]. An infinite number of representations can be constructed from a finite set of primitives, just as the mind can think an infinite number of thoughts and learn new concepts from a seemingly infinite space of possibilities. Structural description models represent visual concepts as compositions of parts and relations. Because the parts and relations are themselves a product of previous learning, their facilitation of the constructions of new models is also an example of learning to learn. For example, new

spoken words can be created through a combination of phonemes or a new gesture or dance move can be created through a combination of more primitive body movements.

Generalization Generalization is the capability of transferring learned behavior about a previous stimulus to novel stimuli. Generalizing newly acquired behavior is an important part of learning and allows an organism to respond quickly and adaptively. Generalization might be based on the perceptual features of stimuli. For example, when a tone (stimulus A) is followed by a shock, conditioned fear will generalize to another tone (stimulus B) to the extent that A and B are perceptually similar. If generalization is based on the perceptual features of stimuli, then it can be said that it is feature based. The second hypothesis is rule-based. Humans can spontaneously create rules, not easily reducible to perceptual features, which allow for efficient generalization of what is learned to novel situations [132]. Take another example : an infant desires milk. An instinct causes it to put various objects in its mouth. A perceptual pattern, which - from an external perspective - we call "mother", repeatedly of whether mother is wearing a tickling sweater or a smooth t-shirt. Once this substitutability is represented as a new entity in the infant's mental system, it has performed an act of abstraction. Such acts of abstraction are often discussed under the term generalization.

Complexity In the context of rules, complexity is different from complexity in information theory. It also different from the concrete / abstract dichotomy evoked above. What we generally evaluate as 'complex', refers to the fact that there are several conditions (hierarchy of cognitive control) and/or a complex processing in time.

1.5.3 Implicit vs Explicit Rules

Rules are explicit constructs but we can learn them either explicitly as in the case of arbitrary symbols such as road signs that are associated with specific meanings or implicitly as in the case of unspoken rules for social interaction.

There has long been a theory of dual processing in the brain, cemented by findings of multiple brain systems that support learning and memory. These memory systems have different operating characteristics, acquire different kinds of knowledge and depend on different brain structures and connections for their operations. Two main strategies are generally reported for the learning and selection of behavior, owing to the two types of functionally disassociable memories : declarative and explicit, or non-declarative and implicit.

In many situations, learning does not proceed in an explicit or goal directed manner. A lot of knowledge and skills are in fact acquired in an incidental and unintentional manner. Implicit or non-declarative memory generally refers to nonconcsious, procedural memory, responsible for skill based kinds of learning. Skill learning, habit formation, simple classical conditioning, priming are all thought to rely on this kind of memory. It is the knowledge that is expressed through performance rather than recollection. It is elaborated by a slow learning processes can generate a rigid behavior, robust in stable worlds, easy to generate but difficult to quickly adapt to changes. Implicit learning is generally presented as sub-symbolic, associative and statistics based. It refers to the non conscious effects that prior information processing may exert on subsequent behavior. It is implemented in associative sensorimotor procedural learning and also in model-free reinforcement learning with biologicical counterparts in the motor and premotor cortex and basal ganglia.
On the other hand, explicit memory manipulating models of the world can be used for the prospective and explicit exploration of possible behaviors, yielding a flexible and rapidly changing strategy, where behavioral rules can be associated to contexts and selected quickly as the evironment or task demand changes. Explicit or declarative memory is the conscious recollections of facts and events. Explicit learning on the other hand is presented as symbolic, declarative and rule based. It is associated with consciousness or awareness, and to the idea of building mental representations that can be used for flexible behavior, involving the prefrontal cortex and hippocampus.

An important point to note here is despite this dichotomy, when one learns to explicitly select a non dominant behavior by cognitive control, repeatedly, it can eventually become a routine, implicit behavior. In other words, the explicit rule can become implicit because it is transferred in the (agranular) regions of concrete rules.

A similar dichotomy or distinction in psychology has long been that between Rules vs Similarity [164]. Rules should eventually, after sufficient training, apply equally well to familiar and novel stimuli. All definitions of the notion of rule take this feature as a defining one. In this context, some investigators believe that learning based on co-occurence statistics between a set of elements, associative learning can only give rise to similarity knowledge, not abstract knowledge. In a network with only implicit representations, knowledge may be represented and activated in response to external stimuli, but it is not available for use by any other part of the system [45]. Only the formation of explicit representations provides a system with

a kind of flexibility and generality such that they can become objects of cognitive manipulation transportable to other tasks.

1.5.4 Hierarchy

The notion of hierarchy has become common in the study of the human cognition in recent time, stating it as a way of how rule represenatations might be implemented in the brain. Another outlook states that cognitive reasoning often involves making hierarchically organized decisions, and also that such decisions then require a causal inference about errors [181], ie reasoning about one's failures by assessing self confidence. Some researchers also believe that there exists a hierarchical organization in motor movements such as reaching (bringing the hand and arm to target). In such accounts, information is processed in a hierarchical system where at the first level, motor related information (for eg, which arm to use) is generated and the target is selected. At the second level, this information is collected and integrated to 'plan' the reaching movement. Finally, at the third level, the planned movement is prepared and executed [98].

There is significant evidence in the neuroscience literature to suggest that the prefrontal cortex is organized hierarchically, and can support such hierarchical processing of information. Precisely how this hierarchy emerges and how it processes and represents information is unclear. There are two common accounts of understanding prefrontal organization - a representation based and a process based account [151]. The first group, as exemplified by Goldman-Rakic, suggests that the functional division of the frontal cortex is on the basis of the different kinds of representational content (eg object vs spatial representations) they encode, while performing essentially the same kind of processing function (eg working memory). On the other hand, the other group suggests that the contribution of different areas is based on different processing functions (eg inhibition vs selective attention [80], or maintenance vs complex processing [160].)

Process based accounts appear to be most useful in the context of incremental extraction of information from the environment, whereby the products of one mental process are used by another mental process. Higher levels exert control over lower levels, for example by controlling the flow of information or by setting the agenda for lower levels. Long term planning problems can be solved using hierarchy because higher levels have a bird's eye view of the problem and can identify appropriate sub-goals while lower levels have a finer temporal resolution and are able to reach each subgoal.

Representation based accounts on the other hand appear to be more useful in the context of attentional and cognitive control processes that require the maintenance of particular information in working memory for the purpose of biasing representations in other regions. Higher levels form abstractions over lower levels, such that lower levels contain concrete, sensory and fine-grained information whereas higher levels contain general, conceptual and integrated information. Cognitive flexibility and task switching can be achieved using a hierarchy over strategies : A high level strategy is in charge of selecting one of several lower level strategies, and low-level strategies guide actual behavior. Whereas the high level strategy is trained to identify the best lower-level strategy for each context, the low-level strategies are trained to optimize behavior within each context.

While the process based and representation based accounts are theories of *what* is organized hierarchically, there are other theories that explain *how* these organizations are implemented. One such theory [151] proposes this dichotomy on a *What* vs *How* axis or ventral vs dorsal pathways. The ventral hierarchy plays a role in guiding the selection and retrieval of semantic/linguistic knowledge while the dorsal hierarchy carries out the processing of the sensory information to guide action outputs. This idea can also be thought of in terms of gradients of *generalization* on the one hand and *complexity* on the other. Generalization refers to having broader categories (eg: color vs red vs brick red) or otherwise being more distantly removed from concrete physical objects (eg: beauty vs sunset). By contrast, complexity refers to the number of different elements that must be taken into account (ie multiple conditionals) to generate a task-appropriate response. For example, the rule "hit left button if the previous stimulus was an A and the current one is an "X" requires two items to be integrated (A and X) to determine the response.

1.5.5 Behavioural evidence of Rules in animals

Initially much of animal learning was thought to be associative or hebbian learning ie hebbian type rules that appear to ensure that jointly activated synapses reinforce each other, and this mechanism may suffice for an ordered separation of pathways and projections, such as those controlling flexion and extension of a muscle. But what remains to be secured is the choice of the correct, adaptive behavioral pattern in response to specific stimulus configuration. This choice could be achieved by a general reward system that would condition the animal to select the correct behavior, but it could also be determined genetically at the structural level. Whereas the

second mechanism may be the norm among invertebrates, it has less commonly been demonstrated in vertebrates.

Foraging and social information seeking are two prototypical types of behaviors displayed by many animals. In natural environments, animals have to face many difficult decision making problems, posed by the details of their habitat and social system. Such problems can include when, where, and for what to forage; with whom to mate and where to nest; whether to flee or to ignore a potential predator. In many primate social groups, males do not mate with all females, because to do so would risk reprisals from dominant males. Yet the presence of a sexually receptive female is among the most potent natural stimuli in the animal's sensory world. That the mating between sensation and behavior is flexible enough to take into account such complex and fluid information as the present state of a group's dominance hierarchy argues against a simple view of stimulus response mappings and for a richer, more nuanced view phrased in terms of decisions.

Most widely replicated test of relational concept learning - same-different task. Another one - relational match-to-sample (RMTS) task, subject must select the choice display in which the perceptual similarity among elements in the display is the same as the peceputal similarity among elemnts in the sample stimulus. For example - AA (sample), subject should select BB rather than CD. Success on such tasks has been reported in chimpanzees, parrots, dolphins, baboons and pigeons.

The ability to make systematic inferences about unobserved transitive relations has been taken as a litmus test of logical relational reasoning. Male pinyon jays can anticipate their own subordinance relation to a stranger after having witnessed the stranger win a series of confrontations with a familiar but dominant conspecific.

Regarding the neural correlates found in the brain, especially the Prefrontal cortex (PFC) and other relatively older brain structures like basal ganglia (BG) and amygdala, the neural processes behind several of the above discussed paradigms in these brain regions will be explored in the following chapter.

In this Chapter, we have introduced and precisely defined some notions and concepts integral to the systematic study of Cognitive Control. While widely studied, Cognitive Control still remains poory understood. In this conceptual overview, we have attempted to tie together the concepts and mechanisms that will be explored through computational modeling in the second part of this manuscript. Particularly, we have begun with the hypothesis that control is a gradient, and that the allocation of control then necessarily depends on the particular cognitive demands evoked by the task at hand. Different conditions will elicit the need for the

different mechanisms discussed. In chapter 4, we illustrate this point through the computational modeling of increasingly complex tasks, thus highlighting the minimal principles and algorithms needed for the implementation of cognitive control. At the first step of this gradient, in two different studies, we illustrate the need of a context to manipulate both innate behaviors, and implicitly acquired behaviors. At the second step, we add a working memory component that can hold the context until the task is completed. Under multiple contexts, we show how an episodic memory component is needed, thus enabling planned and flexible behavior. Finally, at the last and fully elaborated level of Cognitive Control, in chapter 5, we describe the hierarchical elaboration of contexts, which in tandem with the prediction and monitoring of errors, enable adaptive behavior. More generally, the underlying approach in this thesis has been motivated by the development approach of software engineering - to build systems that progressively increase complexity, such that points of failure can then be easily identified, and thus allow for systems that balance complexity and flexibility. We believe that computational neuroscience can benefit from this approach, and could be applied to several questions in the field.

2 Prefrontal Cortex

> Anatomy is not tedious; it is fundamental -
> Richard Passingham

2.1 Introduction

One cannot understand the neural basis of control processes without carefully considering the anatomy of the brain. Understanding the neural mechanisms of control requires delineating specific functional roles of individual neural structures and consequently their functional relationships. The frontal cortex is called so being situated at the anterior end of the brain (anterior to the central sulcus). Comparative studies of the frontal cortex and behavior to understand cognitive abilities can be considered from two perspectives: across the evolution of cognitive abilities over the eons (phylogeny) i.e. across different species; and the development of cognitive abilities during the maturation of an individual member of a species (ontogeny). From this perspective, higher order control over behavior has traditionally been seen as the function of the prefrontal cortex (PFC) [155], which reaches its greatest elaboration and relative size in the primate, especially human brain (Fuster, 1995)[80]. The subcortical and cortical regions work in tandem, creating loops broadly classified under the sensori-motor, limbic and associative labels, with the latter particularly engaged for cognitive control. In humans and non-human primates, these are reported to be in the PFC, which is considered to be the granular part of the frontal cortex [118], i.e., the associative regions and parts of the limbic regions. In rodents, there is no granular part, but some parts of the agranular regions play the role in cognitive control (Box **??**).

Thus, the PFC seems anatomically well situated to play a role in the creation and implementation of abstract rules in primates, and being the hub of cortical processing, be able to exert control over much of the cortex. Moreover, the development of the PFC is also linked to cognitive development, with the rate of brain growth differing for different brain regions [61] [125]. For example, regions that control primary functions, such as motor and sensory systems develop first, followed by the temporal and parietal cortices associated with language and spatial

attention. Of note is that the last brain regions to mature are the prefrontal and lateral temporal cortices involved in the integration of sensory-motor processes, the modulation of attention and critical aspects of decision making and flexible behavior. These cognitive functions are also the latest to develop according to behavioral studies. Much of what we know today about the functional subdivisions in the PFC, comes from lesion and neurophysiological studies in animals, and imaging studies in humans.

The interest in functional properties of the cerebral cortex can be traced back to the early nineteenth century. By that time, it was clear that damage to the cerebrum from war wounds and other causes had a variety of behavioral effects. The anatomist Franz Joseph Gall proposed that differences among individuals in their cognitive functions and personality traits were associated with different parts of the cerebral cortex. His theory gave rise to phrenology, a new approach to studying brain function. Hence, throughout the first half of the nineteenth century, detailed maps were created that classified various aspects of the cortex into categories including memory, color vision, vanity, morality and many others. Today, the concept that distinct areas of the brain are involved in different types of information processing is referred to as *localization of function*. Then came the initial explosion of connectionist research, which held the view of the brain as governed by distributed representations and tuning of synaptic strengths. These fully distributed models became dominant in connectionist psychological modeling. More commonly now, the idea of the brain as an "entangled system" is being promoted [159] where "an entangled system is a deep context-dependent one in which the function of parts (such as a brain region, or a population of cells within a region) must be understood in terms of other parts : an interactionally complex system". An integrated view of the brain, and particularly the prefrontal cortex and how it executes complex cognition may well lie somewhere in between these poles. In any case, some insight about the functional subdivision is quite important to build a comprehensive, well rounded framework for cognitive control.

This chapter provides an overview of the general neuroanatomy of the prefrontal cortex (in humans), its similarities or dissimilarities with other species, and the functional subdivisions, in relation to control processes that have come out after decades of clinical studies. In this section, but also in the wider literature, the focus is primarily on the prefrontal cortex and the basal ganglia. However, other areas are thought to play important roles in cognitive control as well. The pre-supplementary motor area (pre-SMA), inferior frontal junction (IFJ), intraparietal cortex and insula have all been suggested to play an important role in control mechanisms. In the second section, we provide a general overview of the experimental

tasks typically used in the neuroscience literature to probe and test cognitive control.

The PFC across species

Brodmann initially identified a large Regio frontalis in several monkey species [81], including areas occupying the anterior-most lateral, dorsal, medial and orbital surfaces of the hemisphere. These areas together came to be termed as the "PFC", "granular frontal cortex", "frontal association cortex", or a combination of these titles. Brodmann's findings have since sparked a debate since they imply that the granular PFC, and thus the higher level cognitive activities it supports, is missing in the most often used neuroscience model : rodents. Rodents possess homologs of the agranular medial frontal (MF) and agranular orbital areas of primates but lack homologs of the granular cortex that makes up the largest parts of PFC in most primate species. The frontal proisocortex, which is sandwiched between the core isocortex and the periallocortex is comprised of the agranular MF cortex (aMFC), consisting of area 24 (the anterior cingulate area), area 32 (the prelimbic area) and area 25 (the infralimbic area), as well as parts of the orbital and insular cortex adjacent to the isocortex. Primates, but not rodents, have subdivisions of area 32 that are dysgranular as well as agranular, but it is convenient to refer to this cingulate region collectively as aMFC. Similarly, primate orbitofrontal cortex (OFC) is a component of the PFC that includes posterior agranular and dysgranular components, as well as anterior, granular divisions, whereas rodent OFC is exclusively agranular [165]. Furthermore, if the granular PFC is a primate specialization, we should expect it to have features that rodents and other mammals do not have. Such properties, both structural and functional do exist. The granular PFC is part of a wider system of association areas in animals that contain traits specific to primates. Macaque granular PFC neurons, for example, encode correlations between acoustic stimuli and abstract behavior guiding rules [204], as well as associations between color-shape stimuli and abstract problem solving strategies [82]. Lesions of specific parts of the granular PFC in macaques cause profound impairments in rapid learning of arbitrary associations between color-shape stimuli and behavioral goals [38], whereas lesions of the aMFC cause no impairment [37] and even facilitate early stages of learning in these associations in rats performing a similar task [36]. The picture is complex [123], but to draw a simplification, it is reasonable to conclude that rodents have no granular regions, but some agranular regions play the role and act for Cognitive Control. Moreover, the evolutionary addition of granular regions in the primate cortex allows the prospect for contextual control. Further, the difference in the control abilities of human and non-human primates can be attributed to the evolution of the granular cortex itself, with an augmentation in the size of some PFC areas (dlPFC, FPC, anterior granular OFC), without specific new PFC areas per se.

2.2 Cortico - Basal Ganglia loops

There are a number of theoretical (and computational) models that elaborate the role of the basal ganglia in decision making, and action selection (and inhibition). Evidence for the function of the BG in action selection initially came from human patients with damage or dysfunction to this area. For example, Parkinson's disease and Huntington's disease, both cause serious behavioral deficits, ranging from motor (e.g., difficulty in initiating volitional movement) to cognitive (e.g., difficulty in switching tasks).

The main components of the primate basal ganglia are the striatum, the global pallidus (GP) and the subthalamic nucleus (STN) in the forebrain, and the subthalamic nigra (SN) in the midbrain. The globus pallidus contains two subdivisions - the internal and external segments (GPi and GPe respectively), while the substantia nigra contains distinct areas designated compacta (SNc) and reticulata (SNr). In vertebrates, the BG are interconnected with the cerebral cortex and the thalamus. Across species, many aspects of motor function like movements, learning and habituation of actions are believed to be modulated by the processes in BG. Lesions of the striatum, in animal models have shown to produce impairments in learning new operant behaviors (or concrete rules) [72] [163].

A number of computational accounts explain the functional architecture and the possible intrinsic computations in its interaction with cortex and thalamus that drive action selection. Most models of the basal ganglia dissociate two major projection systems : the *direct* pathway from cortex via striatum to the globus pallidus ineterna (Gpi)/substantia nigra (SN), and the *indirect* pathway from cortex via striatum, globus pallidus externa (GPe), subthalamic nucleus (STN) and GPi/SN. These two pathways are thought to interact, by achieving approximately opposite behavioral effects, via opposing effects on BG output nuclei, to produce successful response selection. More recently, there also has been evidence of a *hyperdirect* pathway, implicated in action inhibition, in which the striatum is bypassed, such that the STN forms the input to the basal ganglia.

In general terms, the basal ganglia enables actions by the release of inhibition. In the direct pathway, projections from the striatum release gamma-aminobutyric acid (GABA), thereby inhibiting the downstream neurons in the GPi/SNr, thereby disinhibiting the thalamocortical circuitry to promote movement. In contrast, in the indirect pathway, the striatum counteracts this effect by inhibiting the in-

hibitory GPe. Thus effectively, the SNr firing is increased by the STN-mediated excitation, ultimately inhibiting downstream circuitry and suppressing movement.

The reward circuit is a central component of the network that drives incentive based learning, appropriate responses to stimuli and accurate decision making.

Several parallel and segregated functional loops exist in the basal ganglia. The frontal cortex, the basal ganglia (BG) and the thalamus are associated into several functional loops known as cortex basal ganglia loop (CBG loop). Primarily, based on their functional subdivisions, these are segregated into 3 loops, as described below. However, Alexander et al [1] propose a more precise division in 5 loops.

Different basal ganglia circuits appear to operate predominantly in relation to different types of cognitive and motor actions. These parallel loops, going through distinct regions of the BG, are classified intro three major classes : *limbic, sensori-motor and associative*, shown schematically in Figure 2.1.

- The *sensori-motor* loops (on the left of the figure), involving the dorsolateral striatum, originate in the sensorimotor and premotor cortices, and process exteroceptive information. They are organized around the motor behavior allowing to reach the goal, according to its spatial position (orientation) or according to the physical characteristics involved (handling).

- The *limbic* loops (on the right of the figure) originate in the orbitomedial prefrontal cortex (generally comprised of OFC and the Anterior Cingulate Cortex (ACC)), through amygdala, hypothalamus and the subdivisions of ventral striatum and end back in the medial PFC. Beside processing external information, these loops are based on interoceptive information. They are organized around the selection of the goal of the behavior, according to its motivational value [148], in response to perceived needs or according to its hedonic value [119].

- Finally, the *associative* loops (in the middle of the figure) involve the lateral prefrontal cortex and dorsomedial striatum [1]. Also called cognitive loops, these loops are implied in cognitive control, related to the ability to manipulate abstract rules.

[1] The names of these subregions of the striatum are taken from the literature in rodents. The analagous regions sometimes have different names in primates, but the idea of the subdivision remains the same

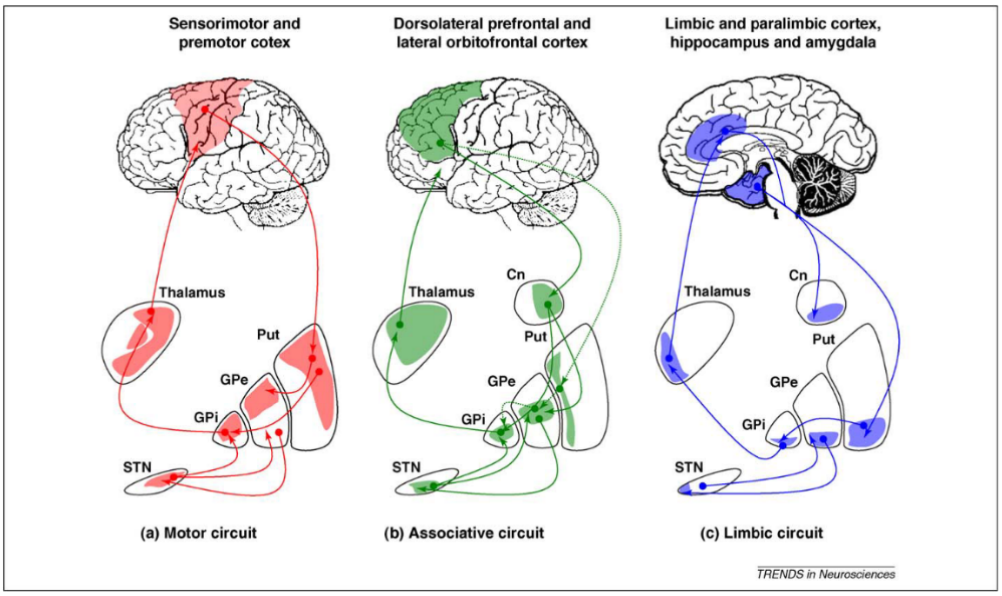


Figure 2.1: Schematic figure illustrating the main cortico-basal ganglia thalamocortical circuits within the human brain. Red : *motor* loops, Green : *associative* loops, Blue : *limbic* loops
(Figure taken from [116])

2.3 Motor/Pre-motor regions

According to the proposed function distinction by Fulton [79], the motor cortex could be divided into a primary motor area (M1, Broadmann area 4) and a premotor area (area 6), which is now further distinguished by its dorsal (PMd) and ventral parts (PMv). Furthermore, the caudal PMd has strong connections with M1 and is well positioned to influence the generation of movements. In contrast, the rostral PMd has strong connections with the prefrontal cortex and selects responses based on arbitrary and spatial cues [42]. In general, when a stimulus is represented in the sensory cortex, candidate actions are generated in the premotor cortices, with both of these regions projecting to the striatum. There also exist oldest agranular and limbic regions, responsible for the simple selection of actions, motivations and goals. The next, more rostral area to these is the pre-PMd (area 8, or the FEF in primates), responsible for attentional processes [162]. This region can also be seen as the most caudal region of the lateral PFC.

2.4 Organization of the prefrontal cortex

The more recent granular regions of the prefrontal cortex are the ones involved in cognitive control. The prefrontal cortex comprises of those parts of the frontal lobe that are anterior to the motor and premotor cortices. Typically, the PFC is subdivided into several functional regions, as illustrated in Figure 2.2. Its lateral surface, as spanned by the inferior, middle, and superior frontal gyri, is usually called **lateral prefrontal cortex**. In some topographical systems, this is further separated into upper and lower parts, termed the **dorsolateral** and **ventrolateral** prefrontal cortex respectively. The ventral surface of the frontal lobes is often called the **orbitofrontal cortex** (i.e. the part of the brain above the orbit of the eyes), and regions along the ventral midline are specifically called **ventromedial** prefrontal cortex. The medial surface of the prefrontal cortex can be roughly divided into anterior and posterior parts, with the posterior, dorsal parts collectively called **dorsomedial** prefrontal cortex. Finally, the most anterior parts of the prefrontal cortex are often called **frontopolar** cortex [166].

Below, we discuss the functional contribution of the medial and lateral parts of the PFC in cognitive control.

2.4.1 Medial PFC

The medial prefrontal cortex (mPFC) is not a homologous unique region, but refers instead to multiple subregions elaborated in parallel. What is referred to as the mPFC in this section, includes the ventral part - ventromedial PFC (vmPFC), the orbital PFC (OFC, the medial aspect of which belongs to vmPFC), the dorsal part - dorsomedial PFC (dmPFC) and finally the anterior cingulate cortex (ACC, which belongs to the dmPFC). The ACC and parts of the OFC are found in many mammals, but they are especially extensive in primates. The idea that these areas of the OFC and brain areas on the medial surface of the frontal cortex, such as the ACC and vmPFC guide decision making, is bolstered by a series of investigations that show that their activity reflects the value of choices [102] [128], the process of decision making [106], the representation of motivation [115] [97], outcome and error [178][177] and the value of the course of action pursued [149] [91]. The medial frontal cortex is more specifically involved in motivating behaviors by monitoring motivationally salient events such as errors, conflict situations, rewards and penalties. In this section, we review the neuroscientific evidence of these claims.

OFC

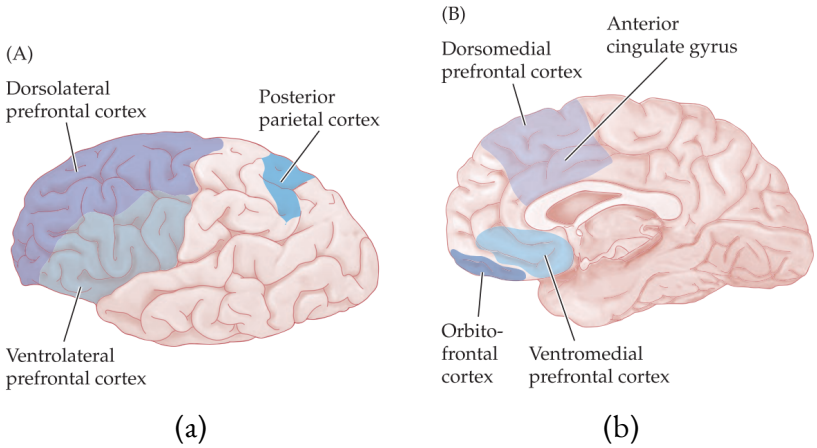


Figure 2.2: **Major brain regions that support executive control**
The brain regions that mediate executive control are (A) lateral structures including the lateral prefrontal cortex (PFC), which is split into dorsal and ventral aspects, and the posterior parietal cortex; (B) midline and inferior structures including the ventromedial PFC, orbitofrontal cortex (which continues to the ventral surface of the frontal lobes, not shown), and dorsomedial PFC, which includes the anterior cingulate gyrus
(Figure taken from [166])

Lesion studies in rats and monkeys have consistently implicated the OFC in the guidance of flexible behavior. There are two consistent impairments found in such lesion studies : perseveration of previously rewarded choices following the reversal of deterministic stimulus-outcome associations [43] [101], and insensitivity to outcome devaluation [23] [18]. One proposal that followed these results was that there seemed to be an insensitivity to negative reinforcement (errors) [73], or an inability to inhibit previously rewarded actions [46]. More recently, studies [175] [185] have shown that monkeys with OFC lesions performing 3-alternate choice tasks switch as frequently as controls, rendering the account unlikely. An alternative proposal is that the OFC plays a role in representing outcome expectations predicted by cues in the environment [176]. Generally, animals with OFC lesions fail to reduce responses to devalued cues (ie when previously rewarded cues are 'devalued'). This suggests that the OFC may contain information about the subjective value of specific outcomes. However, owing to its connectivity with several other specialized brain regions, pinning down a specific and unique functional role has proved difficult. In general, it is implied in affective decision making.

vmPFC and dmPFC

Decision neuroscience studies have long suggested a disassociation between the ventral and dorsal medial prefrontal cortex (vmPFC and dmPFC), which provide estimates of goal value and action cost, respectively. Everyday life often requires arbitrating between pursuing an ongoing action plan by possibly adjusting it versus exploring a new action plan instead, or deciding to either choose between currently available options or foraging for other, possibly better ones. Resolving this so called *exploitation exploration* dilemma involves the mPFC, and shows a difference in the functional role of the ventral and dorsal parts.

In typical studies attempting to dissociate the functional role of these parts, choice or preference tasks are performed. A standard view emerged through these studies [114], that during decision making, vmPFC/mOFC acts as a choice option comparator. It represents potential choice options, computes their comparison, and turns them into actual choices in the frame of reference currently relevant for guiding actions [88]. It reflects the relative evidence for taking one choice over another. Activity in the vmPFC has also been shown [110] to correlate with the expected monetary value (reward probablility x reward magnitude) associated with a stimulus even in the absence of choice. Extensive evidence also indicates that vmPFC activity scales with the value of the chosen option during decision making [22] [74] [99] .

In the study by [44], FMRI results suggest the role of the vmPFC as a generic valuation system, its activity increasing with reward value and decreasing with effort cost. In contrast, more dorsal regions are not concerned with attributes of options but with metacognitive estimates, confidence level being computed in mPFC and deliberation time in the dmPFC.

Some interpretations insist that the dual role of the vm/dmPFC is based on the estimation of costs and benefits [158], implicating the vmPFC in the estimation of reward values and the dmPFC in the estimation of cost effort [17] [28]. A related view is that the vmPFC signals values in a space of goods, whereas the dmPFC encodes values in a space of actions. While this view is supported by empirical evidence in a number of studies [154] , other researchers have found effort cost representations in vmPFC activity [131] [208] and reward value in dmPFC activity [109]. Other accounts are based on the comparison between options during choice, suggesting that both regions estimate decision values, but in an opposite manner. For example, for value difference, the vmPFC activates while the dmPFC deactivates. Closely linked to this view is the account of vmPFC activity signalling the value of the default or dominant behavior (in other words, the reliability of the current strategy, or exploitation), and the dmPFC activity signalling the value of alterna-

tive options (the need for exploration / switching) [114]. Human neuroimaging studies have shown that activation in the vmPFC reflects the subjective value of the ongoing plan according to action outcomes, whereas the dmPFC exhibits activation when this value drops and the plan is abandoned for exploring new ones [64].

This idea was elaborated in a study done by Domenech et al [63]. The authors discovered that the neural activity in the vmPFC infers and tracks the reliability of the ongoing action plan in order to proactively encode the upcoming action outcomes as either learning signals or potential triggers to explore new plans. By contrast, they claim that the dmPFC exhibits neural responses to action outcomes, which results in either improving or abandoning the ongoing plan. Thus, their account suggests that the mPFC solves the exploitation - exploration dilemma by a two stage predictive process : the first being the proactive ventromedial stage, followed by a reactive dorsomedial stage.

In summary, a key question to ask is : what precise computational role does the vmPFC play during RL and value based choice ? Four non mutually exclusive proposals have been emerging : the vmPFC *1* compares the evidence between competing options, *2* provides input to a decision making integrator elsewhere in the brain, *3* encodes the subjective value of the chosen option and *4* transmits the chosen value predictions to DA neurons for prediction error computation

ACC

The anterior cingulate cortex (ACC) commonly refers to the cytoarchitectonic areas 24 and 32. In the neuroscience literature of cognitive control, the dorsal ACC (dACC) has an almost ubiquitous presence. It has been implicated in a diversity of functions, from reward processing and performance monitoring, to the execution of control and action selection.

Converging experimental evidence shows that dACC activity (1) encodes the distribution of opportunities across time as well as space, (2) assesses the value of disengaging from the present course of action and (3) regulates switching between periods of exploiting such knowledge and seeking more information [108]. Both neuroimaging studies in humans and single neuron recording studies in macaques demonstrate that the dACC simultaneously holds multiple representations of value with different time constants. When lesions are made in the ACC, macaques can only adjust their behavior in response to the most recent outcome, but the influence of the longer history of reward and choice is lost. Neuroimaging studies also confirm that the dACC has a preeminent role in information seeking; its activity

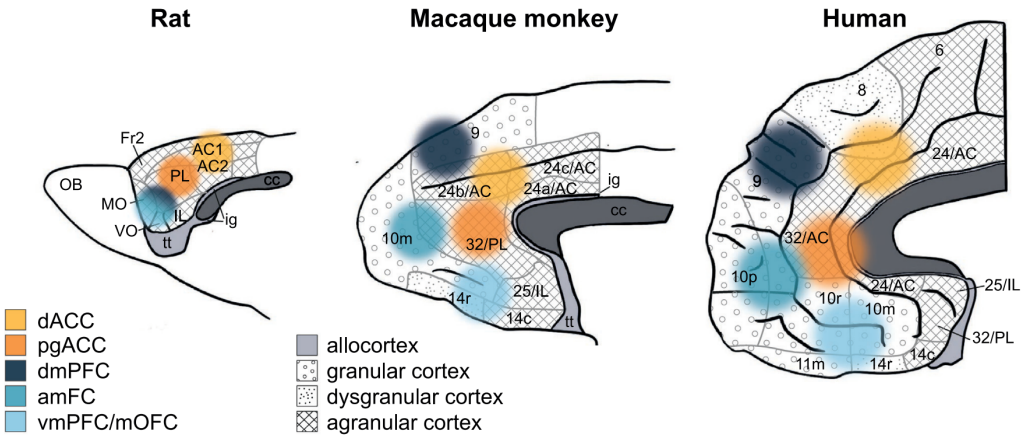


Figure 2.3: **Medial and orbital frontal cortex in rodents, macaques, and humans** Five functional regions, dorsal anterior cingulate cortex (dACC), perigenual anterior cingulate cortex (pgACC), dorsomedial prefrontal cortex (dmPFC), anterior medial frontal cortex (amPFC), and ventromedial prefrontal cortex (vmPFC/mOFC) are shown in relation to cytoarchitectonic maps of rat (left), macaque (center), and human (right) medial and orbital frontal cortex. The color scheme indicates that the orbital region of rodents has some functional features shared with primate dmPFC, amFC, and vmPFC/mOFC but that it does not correspond in a simple way to any of them. Thus, while the extent to which regions are homologous across humans and other primates, such as macaques, is relatively clear, correspondences between primates and rodents are more contentious and unlikely to be one-to-one in nature. (as illustrated in [108])

reflects a person's uncertainty about the choice that they are taking when they are actively exploring options to obtain information rather than when they are simply randomly responding. In sum, the dACC represents the distribution of opportunities in the environment, it computes recent and long term value, and on these bases, it determines whether a person or other animal should engage with a current option or explore the environment, including driving specific information seeking activity.

2.4.2 Lateral PFC

The lateral part of the prefrontal cortex is engaged in cognitive control, related to the ability of the PFC to manipulate abstract rules when the selection criteria is required to be more elaborated. Especially dorsolateral prefrontal cortex (dlPFC) in

primates has been found crucial for the most flexible, complex, and expectation-oriented behaviors that need to be organized, planned and produced [139]. Lateral PFC is connected to wide range of secondary sensory regions like Frontal Eye Fields (FEF), secondary visual cortex, parietal cortex, supplementary motor cortex and pre-motor cortex. When the selection is not trivial and requires memory, context, and abstract rules combining them, lPFC complements additionally the other PFC systems and the downstream selection mechanisms, and hence seems to be a critical area for the representation of rules [143] . Patients with lateral PFC damage report correctly on what the appropriate task rule is, even while being unable to implement it correctly [85]. The developmental literature makes similar observations, suggesting that the growth of knowledge sometimes proceeds faster than the ability to control behavior.

Conventionally, the lateral prefrontal cortex is divided into the dorsolateral prefrontal cortex (dlPFC) and the ventro lateral prefrontal cortex (vlPFC). These two areas are separated by the inferior frontal sulcus.

The dorsolateral prefrontal cortex (dlPFC), along with the posterior parietal cortex (PPC) has been implied in the encoding of rules. Functional neuroimaging studies of action-rule switching typically report activation in a distributed network encompassing the dlPFC, vlPFC, premotor cortex (PMC), pre-SMA, and the PPC. Studies show that individual neurons in the region fire categorically different responses depending on the specific or abstract rule used [203][146] [78]. It has also been shown that this region plays an important role in working memory [35] [140].

The dlPFC has long been considered to house representations that guide temporally integrated, goal-directed behavior. [161] [213]. Recent work has refined this idea by demonstrating that dlPFC neurons play a direct role in representing task sets. [10] [33] [209]

The ventrolateral prefrontal cortex (vlPFC) is viewed to be more involved with the visualization with more precision, possibly owing to its connections with the sensory areas in the inferotemporal cortex and the auditory temporal gyrus. This region has also been implicated in rule learning [33], retrieval form long term memory and on line maintenance during task preparation. Although vlPFC has been regarded to play a role in learning stimulus outcome associations as does the orbitofrontal cortex (OFC), an interesting disassociation has been pointed out that OFC is necessary for updating associations that signal desirability whereas the vlPFC is necessary for updating associations that signal availability [142]. The vlPFC also has outputs to dlPFC, and taken together, the findings suggest that it is possible that vlPFC defines a goal, and governs the processes in the dlPFC, transforming the information from stimulus to behavior. Or on the contrary, for more complex

and planned behaviors that have subgoals, then the rules in the dlPFC activate the vlPFC which maintains subgoals.

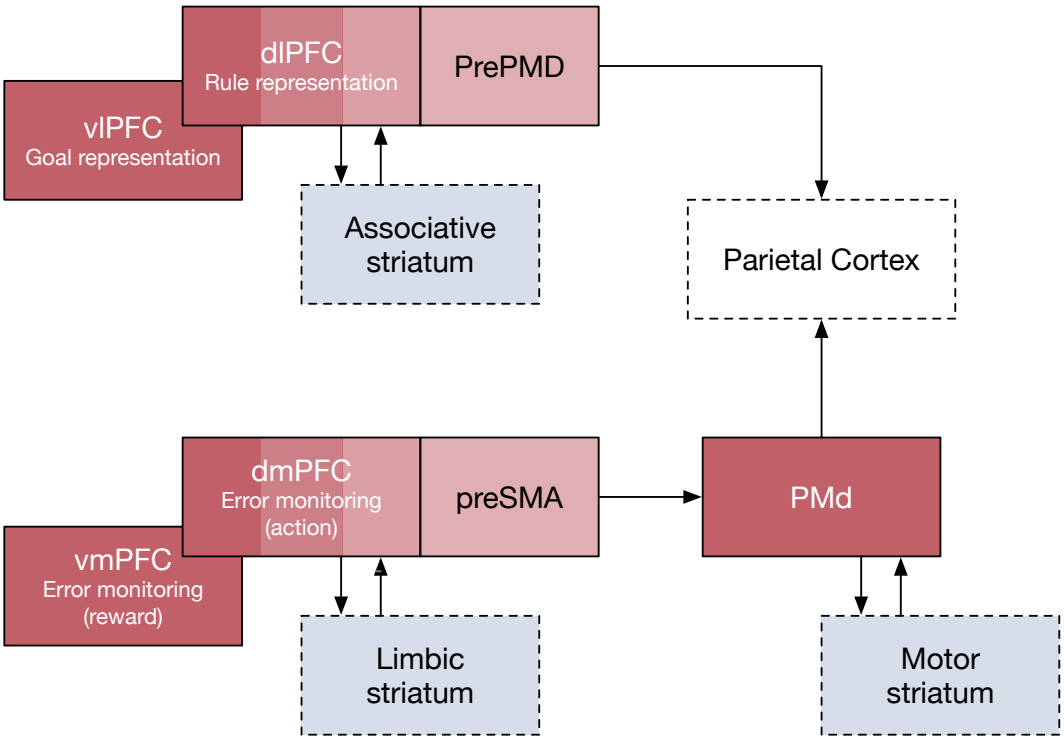
2.4.3 Hierarchical organization

According to the research discussed thus far, the lateral PFC plays a critical role in cognitive control. Left unresolved are questions pertaining to the wider functional organization of the PFC, i.e., are the different functional roles ascribed to different subregions organized into any sort of topography ? For a long time, neurologists have recognized that the more posterior regions of the frontal lobes are more closely linked to motor activity, while the more anterior regions support processes related to reasoning and mental simulation [15]. Furthermore, the posterior frontal regions mature relatively early during human development and also share more similarities with nonhuman primates, whereas anterior regions like the frontopolar cortex develop late. Therefore, it makes sense to hypothesize that the PFC is organized in rostral-caudal manner locally (i.e., within the dlPFC or dmPFC), in accordance with the increasing abstraction of cognitive control, with anterior regions supporting complex functions related to higher-order behavioral goals and posterior regions supporting simple functions linked to matching behavior to stimuli. As discussed in Section 1.5.4, two main theories have been proposed.

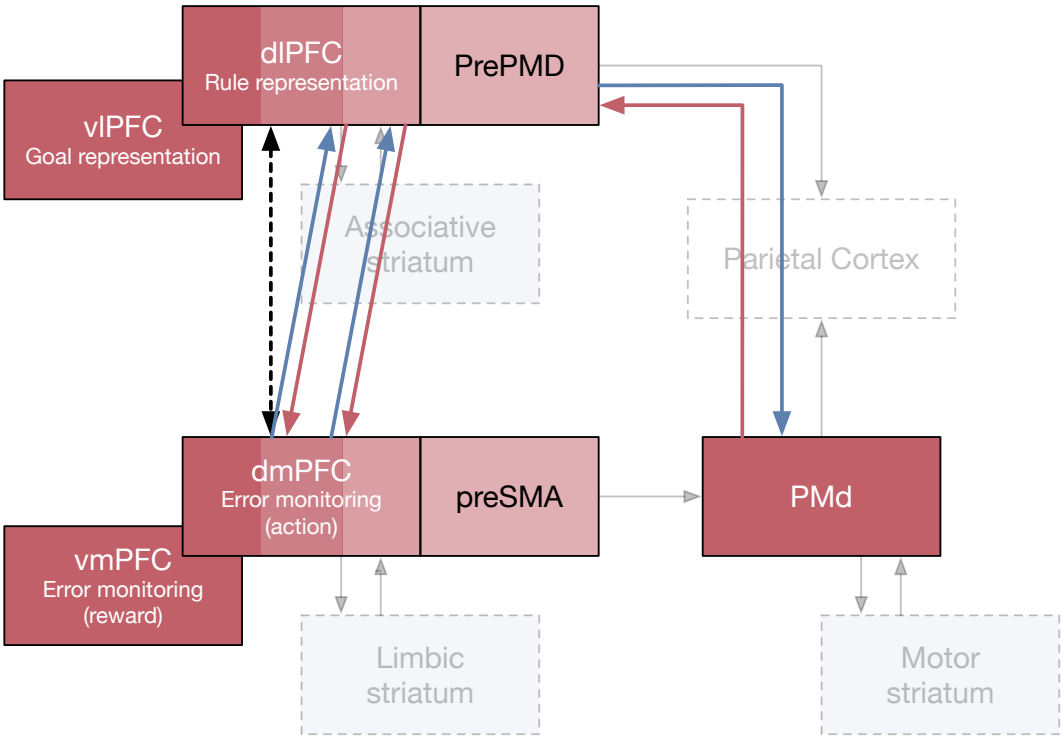
The first theory by Koechlin, Ody, and Kouneiher [112] [113] contends that executive functions are organized according to the degree of temporal abstraction. The lateral PFC is involved in cognitive control by forming a hierarchy of top down selection processes from posterior to anterior regions for selecting appropriate behaviors, according to the temporal structure of events involved in action selection, which defines the crucial levels of cognitive control. Specifically, sensory control involved in selecting motor actions in response to currently observed stimuli is subserved by lateral premotor regions. Contextual control is subserved by more caudal lPFC regions which is involved in selecting S-R associations depending on contextual signals that accompany stimulus occurrences. Finally, episodic control is subserved by rostral lPFC regions involved in selecting caudal lPFC representations (task sets) that relate to ongoing internal goals for behavior. Posterior lPCF regions subserve transient control by selecting sensori motor associations for immediate action according to information conveyed by concomitant contextual signals. In contrast, middle lPFC regions subserve sustained control over behavioral episodes by adjusting selection in posterior lPCF regions according to information conveyed by temporally remote events [115]. Further, these authors propose a similar hierarchical organization in the mPFC.

According to the other theory, the authors Badre, Hoffman, Cooney, and D'esposito [12] [11] propose policy abstraction as a substitute for traditional organizational structures. They concur with Koechlin et al, that posterior regions facilitate the creation and application of straightforward rules that connect stimuli to behavior (such as "Click the button when you see a red square"). However, they argue that the more anterior regions support higher-order behavioral policies, which are required to choose which of a number of straightforward rules applies in the given situation. Their research suggests that executive functions are organized in a hierarchy, such that more complex processing of the anterior regions shapes the functioning of the posterior regions. They do this by incorporating several levels of increasingly complex rules into a single experiment, ranging from simple stimulus-response mapping to meta-rules that govern how a task should be approached.

The question then is whether there is an overarching principle of functional organization for the PFC, engaged and responsible for cognitive control. Based on the literature reviewed in this section, we propose a functional organization as illustrated in Figure 2.5. We make a distinction between rule learning on the one hand, and rule representation and implementation on the other.


dlPFC
Rule representation
PrePMD
vlPFC
Goal representation
Associative striatum
Parietal Cortex
dmPFC
Error monitoring (action)
preSMA
PMd
vmPFC
Error monitoring (reward)
Limbic striatum
Motor striatum

(a)


dlPFC
Rule representation
PrePMD
vlPFC
Goal representation
Associative striatum
Parietal Cortex
dmPFC
Error monitoring (action)
preSMA
PMd
vmPFC
Error monitoring (reward)
Limbic striatum
Motor striatum

(b)

Figure 2.5: **Functional organization of PFC**

(a) Rule Learning : Sensory representations are encoded in the posterior cortex, with two kinds of information : *what* in the posterior temporal cortex and *where/how* in the posterior parietal cortex; the motor-premotor cortex encodes concrete sensorimotor rules, and transforms sensory representations into actions; in the limbic regions of the frontal cortex, values of goals (vmPFC) and values/cost of actions (dmPFC) are learned, along with interactions with the striatal regions. These learned values can then be used to select the best concrete rule from the situation (stimulus driven) or depending on a current goal (goal driven). Typically, the same rule is selected in certain contexts (default or habitual behavior).

(b) Rule representation and implementation : There is another region of the dmPFC (same for vmPFC), which can detect conflicts (several close possible solutions), predict outcomes, and compute errors of prediction; this region can inhibit the default behavior and explore to trigger another, more adapted concrete rule. This is the basis of cognitive control. When the contextual condition to select another concrete rule is too complex, an abstract rule must be created to represent it. This is done in lateral PFC, to pay more attention to some cues in some context, with vlPFC to define such complex hierarchical cues (goals and subgoals) and dlPFC to create a sequence of actions to get these cues (strategies or abstract rules). This lateral region of PFC in loop with associative region of striatum, can learn internal actions of WM monitoring (input and output gating); Abstract rules can be hierarchical, both in lPFC and mPFC. In lPFC, the lower level of rule corresponds to PrePMd (FEF in primates) for attention monitoring. In mPFC, this difference of level is between PreSMA (conflict monitoring at the lower level) and dACC (error monitoring at higher levels)

2.5 Experimental Tasks

Cognitive control is studied using a wide variety of experimental paradigms that measure switch costs, perseveration costs, anti-saccade latencies, percent recall, and stimulus response compatibility effects.

2.5.1 Working Memory paradigms

n-back : Paradigms that test working memory typically require processing the properties of current stimulus and making response related decisions, while simultaneously maintaining and updating information held in working memory. An example is the N-back task, in which participants are shown a series of stimuli (words, numbers) and are required to respond to the current stimulus if it is identical to a stimulus presented one, two or more (n-back) trials ago. Cognitive control load is manipulated by increasing the length of the sequence that needs to be maintained in working memory.

Matching and non matching rules : Delayed matching to sample (DMS) and delayed non-matching to sample (DNMS) tasks require rule-based comparisons of the sameness or difference between stimuli that can be generalized to multiple exemplars, including novel items. Generally, monkeys are trained to apply DMS and DNMS rules to visual items with a familiar cue at the start of each trial indicating the relevant rule to be applied. Monkeys need to be able to maintain the rule information across a delay period to be able to apply the rule when the sample and test items are shown.

2.5.2 Conflict paradigms

These paradigms introduce interference that results in a conflict between task relevant and task irrelevant stimulus properties and/or stimulus - response associations. Cognitive control is engaged to monitor this conflict. The conflict manipulation is achieved by presenting stimuli that contain features associated with different responses.

Eriksen flanker task : This task requires participants to respond to the central stimulus in an array, which is flanked by non-target stimuli which correspond to either the same directional response as the target (congruent flankers, or compat-

ible condition), to the opposite response (incongruent flankers, or incompatible condition), or to neither (neutral flankers). Subjects need to suppress responses that inappropriate given the context, and hence the task measures the ability of the subjects to engage cognitive control in the inhibition of competing responses.

Stroop task : In this task, participants must attend to one dimension of the presented stimulus (eg the color in which the word is displayed) and ignore a competing but prepotent dimension (the word itself). This task thus tests the capacity to inhibit the irrelevant or interfering stimulus or response representation.

Hierarchical discrimination tasks : In the hierarchical discrimination tasks reported by [13], [112], there are typically 2 or more first order stimulus response associations based on specific dimensions of the stimuli (for e.g., vowel/consonant discrimination and upper/lower case discrimination) and the conflict between these lower order responses is resolved based on another dimension of the stimulus (eg, color - red means task vowel/consonant, blue means task upper/lower case)

2.5.3 Response inhibition paradigms

These paradigms involve monitoring conflict between a prepotent tendency to emit a response and the need to withhold that response under certain circumstances.

Anti-saccade task : This task requires the subjects to stop a reflexive eye movement to a brief cue and instead look voluntarily in the opposite direction to identify a briefly appearing target stimulus before it is masked.

Go/NoGo task : Generally, this kind of task involves a primary task that requires a response (eg, a 2-choice decision task such as "press left for X and right for O") as well as an infrequent contingency condition that requires this response to be withheld (eg, do not respond to red stimuli). The properties of the stimulus determine whether to implement the S-R association, select and execute a response or alternatively, to interrupt this process and wait for the next trial.

AXCPT and 12AX task : These are both continuous performance tasks which mix elements of working memory tasks with those of go - nogo and hierarchical tasks and therefore tax multiple aspects of cognitive control. In the AXCPT, a response needs to be given to X on a trial, but only if X was preceded by an A on the immediately preceding trial. The elaborated, conditional 12AX task stresses the sequential interaction between the contents of the working memory and direct stimulus input. The task is to respond only for X of "AX" in the case that the most recent number has been a "1" and the Y of "BY" if the most recent number had been a "2".

2.5.4 Task switching paradigms

Wisconsin Card Sorting Test (WCST) : The most established paradigm that involves non-cued selection of abstract rules and flexible shifting between such rules is the WCST, routinely used in neuropsychological assessments to test cognitive flexibility. In the WCST, the relevant rule for matching (color, shape and number matching) is changed without explicitly cueing when correct performance under the current rule meets a predetermined criterion, after which participants need to discover the new rule by trial and error. The key measure that indicates the ability to switch quickly are the number of perseverative errors the participant displays (the tendency to stick to the current rule). Children typically persevere more than adults.

Intra and extra dimensional shifting and reversal learning task : Select shape or line stimuli based on rules that are inferred based on feedback. After some number of correct trials, the rule changes. Shift trials require applying old rules in new stimuli in the same shape category (intradimensional) or the other line category (extradimensional). Reversals require selecting the previously ignored stimuli within or between categories.

Task-Set Switching : Subjects are typically required to switch between two (or more) tasks. For example, when two tasks, task A and B need to be performed, the sequence of tasks may switch from one to the other at random : AAABBBAA. Task-set switching is the process of selecting between the two competing task sets.

In this chapter, in light of the evidence from neuroscience, we have proposed a functional organization of the PFC that supports cognitive control. In the second section, we have identified commonly used experimental paradigms that are used to test it. This functional description will form the basis of our computational modeling efforts (Part II of the manuscript). In the next chapter, we present the current state of the art in computational modeling, and discuss the extent to which the existing models explain or contradict the evidence from literature. Further, we also place an emphasis on models that take into account the architectural or mechanistic constraints that are highlighted by the literature reviewed here.

3 Computational Models

A number of computational models have played a prominent role in the development and understanding of cognitive control and its underlying mechanisms. These models can be roughly partitioned into two categories : computational models and cognitive models. Cognitive models generally provide a higher level abstraction using symbolic approaches that may provide some hints but in the meantime, they rarely explain the neural basis and for this reason, we will discard them in the present review. On the computational side, there are also a very large number of models that provide explanation and implementation of different aspects of cognitive control. Since the number of such models is also very large, we will concentrate in this review on the most comprehensive and contemporary models of cognitive control, which have focused on some key functions integral to cognitive control : namely working memory, monitoring and top-down control. Computationally, these translate into three core challenges :

- *How does the brain determine what information is relevant to be maintained during an ongoing task goal, and when this information should be updated with newer information ?*
- *How is the current demand for control evaluated and what is the necessary relevant information that underlies this neural computation ?*
- *How can higher level goals constrain and implement a lower level goal (or rule)*

In the following sections, we review existing models that model each of these functions, and the challenges associated with each of them. A summary of these models is listed in table 3.1.

3.1 Working memory models

As defined in the section 1.4.5, working memory refers to that process which allows an agent to store, update and retrieve information that may be relevant for an ongoing task, and thus supports abstract, goal directed behavior. Beyond the question of the neural mechanisms responsible for the storage/update and retrieval of

this memory, there are also questions related to the content and timing. More precisely,

- Which information needs to be stored ?
- When does the information need to be stored ?
- When does this information need to be updated ?
- When does this information need to be retrieved

A number of models have attempted to simulate this process, in a biologically plausible manner. These models attempt to put forward a computational theory that answers the question,

Early neurocomputational models [40] [216] [50] [60] used attractor models to investigate the mechanisms through which working memory could be actively maintained against irrelevant distractors. These models characterize how stable, but flexible representations can occur in biologically plausible neural networks (for eg, some forms of *recurrent neural networks*), which lead to stable attractor states, resembling the brain's neural activity during working memory maintenance. However, most of these models lacked a mechanism for updating the working memory when newer, task relevant information was presented. In fact, this contradiction between the two main working memory functions - that of active maintenance vs flexible updating - is difficult to reconcile, since maintenance increases the resistance to distractors, while updating makes the system vulnerable to them. Thus, this problem poses a fundamental computational problem in learning when to maintain context representations and when to rapidly update information into working memory, which is also a core problem of cognitive control.

The most well established model for working memory that was able to present a solution to the problem, from the connectionist or neural network tradition, is the prefrontal cortex and basal ganglia working memory (PBWM) model developed by O'Reilly and Frank [150]. In this model, the prefrontal cortex (PFC) and basal ganglia (BG) interact to solve the maintenance versus updating problem by implementing a flexible working memory system with an adaptive gating mechanism. The working memory is insulated from irrelevant sensory input when the gating mechanism is closed, and is receptive to updating information when the gating mechanism is open. At the biological level, the model proposes that the PFC facilitates the maintenance of information while the BG performs the dynamic gating via disinhibition. Dopaminergic "Go" neurons in dorsal striatum fire to disinhibit

PFC to enable updating of working memory representations, while "NoGo" neurons support maintenance of PFC working memory representations.

Another model Todd, Niv, and Cohen [197] provides an abstraction of this model by framing the concept of working memory as an internal (hidden) state representation, using a partially observable markov decision problem (POMDP) framework. The model learns *when* to update the internal working memory element by online temporal difference methods, by using a tabular version of the RL algorithm.

Yet another, recently proposed model, Working Memory Through Attentional Tagging (WorkMATe) [120] combines the gating mechanism insight in LSTM and PBWM models, takes a simple, biologically plausible learning algorithm of the AuGMEnt model [172], and uses abstract stimulus representations. This model takes longer to train on tasks compared to other models, but is able to generalize across previously unobserved stimuli as compared to the simplified PBWM by Todd et al.

Although these models capture certain aspects of reinforcement learning and working memory in a biologically realistic way, they typically need very long timescales to train, and also do not capture the resource limited capacity of working memory in both humans and animals.

3.2 Models of monitoring

Several theories regarding the interpretation of the ACC's computational role in cognitive control have implicated the detection of error signals, conflict monitoring, error likelihood detection and calculating uncertainty as likely candidates.

A prominent example of the framework of control is the conflict model proposed by Botvinick, Braver, Barch, Carter, and Cohen [27] . The *conflict monitoring model* identified the role of ACC as a conflict monitor that increases in activation as a function of conflict between available responses. According to this account, incongruent stimuli, ie stimuli whose response is incompatible on two (or more) dimensions (eg, word reading and ink color in the Stroop task (section 2.5.2)) will activate competing responses. Such competition leads to a conflict in response selection, where conflict is defined as the multiple activity of the competing channels. The model proposes that the ACC resolves this conflict by tracking the evidence for a need to increase cognitive control, and sends this information to the dlPFC. The dlPFC then exerts control over the processing in posterior brain areas, and favors one solution with an attentional process. However, lesion studies have

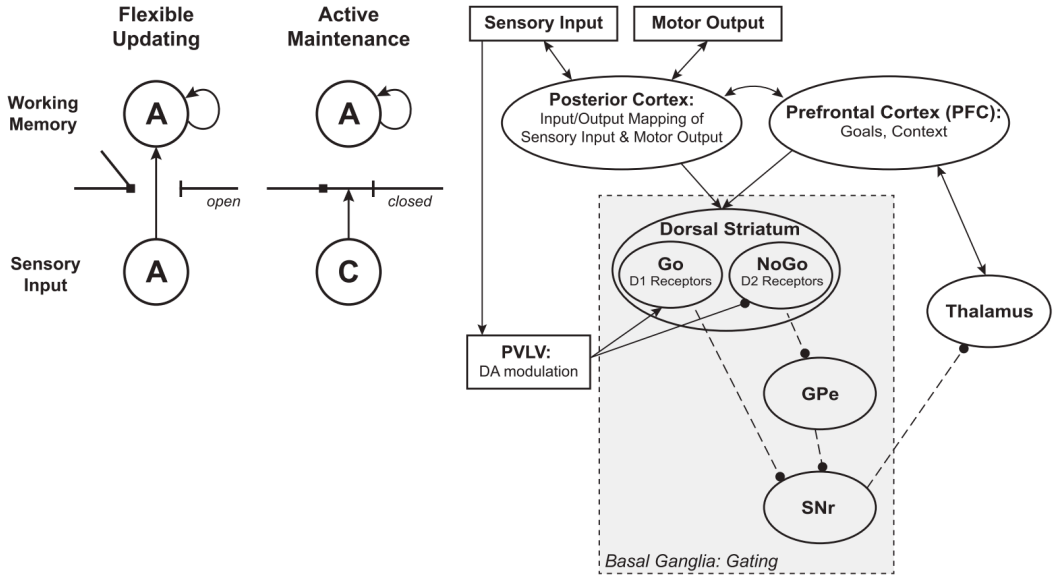


Figure 3.1: **Frank and O'Reilly's Prefrontal Basal Ganglia and Working Memory (PBWM) model (a)** Gating mechanism which switches between active maintenance and flexible updating of working memory in incorporate task-relevant information **(b)** Neural Network model implementation of PBWM.
Figure taken from [150]

shown that ACC lesions do not always impair the cognitive control adjustments that are likely to follow after conflict detection, according to this theory.

A related model was later proposed by Brown and Braver [32], called the *error likelihood model*. They also posited that the ACC activity regulates the activity of other structures involved in implementing cognitive control. This account proposed that ACC associates errors to stimulus-context in which they occur, thus effectively trying to explain both the conflict and error activity. While subsequent experiments have verified the critical aspects of this model, it was unable to simulate the effects of unexpected errors, ie errors that are committed in contexts with low error likelihood.

A rather different model has been proposed by Holroyd and Coles [95], [94], ascribing a role in *action selection* to the ACC, shifting it from the evaluation domain role. In this view, the ACC acts as a "motor control filter" with action policies learned via a dopaminergic teaching signal (reward prediction error). There is evidence for ACC encoding action values in uncertain environments; however this model only explains a mechanism for action selection, but does not make specific predictions about how these reward and error signals can regulate behavior.

The theory put forward by Behrens, Woolrich, Walton, and Rushworth [19] proposed that the ACC is sensitive to the *volatility* of environmental outcomes. According to this theory, the ACC is responsible for detecting how rapidly reward contingencies in the environment are changing over time. The model puts forth a mechanism for flexibly adapting the rate at which the current knowledge of the world is updated with new information : the volatility measure computed by the ACC is used to dynamically adjust this learning rate. What the model also insists on, is that the volatility signal is separable from the prediction error signals (mentioned in the previous model) and that these two can coexist within the ACC. On the other hand, this model does not explain the contribution of the ACC to action selection.

Recent studies have started to investigate the role of the ACC in processes that require effort, showing that the ACC activity increases when subjects prepare for difficult or effortful tasks, even without the presence of conflict, error or choice. For example, ACC lesions have been shown to be associated with motivational impairments and apathy in humans, and in non human animals, they impair decisions that evaluate trade-offs between effort and reward value. These findings, led to the proposal that the ACC might encode for *choice difficulty* (conflict between choice options). The *adaptive effort allocation model* by Verguts, Vassena, and Silvetti [202] proposes a "boosting" mechanism implemented by the ACC, which is able to bias behavior towards effortful options when they are predicted to return a large enough reward. In line with this conception of the ACC role, the *expected*

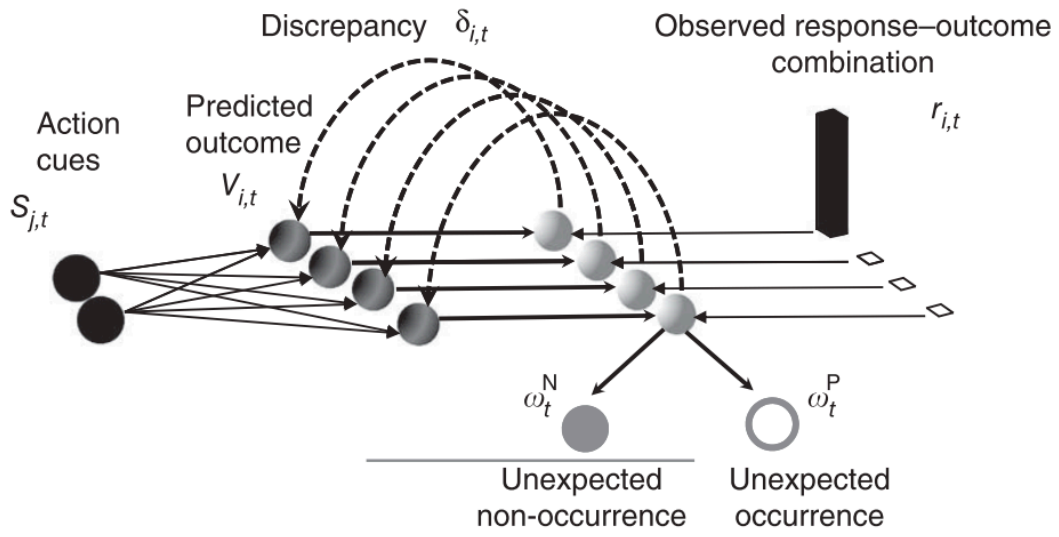


Figure 3.2: **Alexander and Brown's Prediction-Response Outcome (PRO) model**
The model learns predictions of multiple possible outcomes of various chosen actions using an error likelihood signal.
Figure taken from [4]

value of control (EVC) framework [187] posited that the ACC computes a "value of control" by integrating a variety of the signals mentioned above.

A more recent model [4] called PRO (prediction response outcome) provides a reconciliation to several of the competing accounts of mPFC function mentioned above. The PRO model assumes the medial PFC learns a forward model predicting multiple likely outcomes of chosen actions, and tracks the discrepancies between actual and predicted or expected outcomes, and uses this error signal to refine future predictions. Interestingly, the prediction error signal in this model also signals a "negative surprise", when an expected outcome does not occur. Using these disassociated signals, the model posits that the prediction signal increases reliably immediately prior to when the most likely outcome will occur, while the negative surprise signal activates reliably after the action that produces the unexpected outcome has occurred. The model is able to explain most empirical findings (for both neuroimaging and single cell data) supported by previous account of the mPFC (conflict, error, reward prediction errors), making it a useful generalisable computational framework for ACC's role in cognitive control.

3.3 Models of top-down control

The notion of applying hierarchical structure to parse complex systems into subordinate and interrelated systems has long been established, with subsystems being further subdivided into 'elementary' units [93].

The *information cascade* model proposed by Koechlin, Ody, and Kouneiher [112] argues that control signals used to guide behavioral actions, based on internal plans and goals, can be subdivided into sensorimotor, contextual and episodic control. According to this model, the division of control is based on the temporal aspect of when control is implemented, with actions selected based on temporally proximal stimulus lower on the hierarchy and actions selected based on past information higher on the hierarchy. At the neural level, this model gives an account of hierarchical organization along the anterior-posterior axis of the lateral PFC for storing hierarchical rules, with more hierarchically superior control signals being represented in the more anterior cortical regions. Kouneiher, Charron, and Koechlin [115] have shown that there is a similar anterior-posterior division in the medial PFC for control.

Ribas-Fernandes, Solway, Diuk, McGuire, Barto, Niv, and Botvinick [170]'s hierarchical model extends standard RL approaches to address the higher level organization of behavior in lateral PFC. According to this framework, RL principles operate simultaneously at the levels of temporally extended behavioral sequences (options) and the lower level actions that make up those sequences. A central claim of this theory is that achieving the subgoal specifically by a particular option should act as a reinforcing event in the absence of explicit reward. Effectively, it is argued that the PFC is able to hijack the brain's basic reward mechanisms to reinforce behaviors consistent with the organisms' high level goals. This hierarchical framework provides a formal model of the emergence of high level task structure in lateral PFC through reinforcement learning, but does not directly identify the specific contributions of medial PFC in the RL process.

This idea of hierarchical organization of action sequences has also been implemented by the *HRL-ACC* model by Holroyd and McClure [96], to account for how the ACC selects and motivates the execution of extended behaviors, rather than encoding moment to moment changes in behavior following conflicts and errors. The idea was implemented using a three level hierarchical reinforcement learning model, to explain rodent behavior. In particular, the 'high level' module (rostral ACC) selects the 'meta task' for execution, which applies a control signal to the middle (caudal ACC) module based on effort cost. This module then selects the task for execution and further applies a control signal to the lower module which implements the action selection mechanism. While on the one hand, this

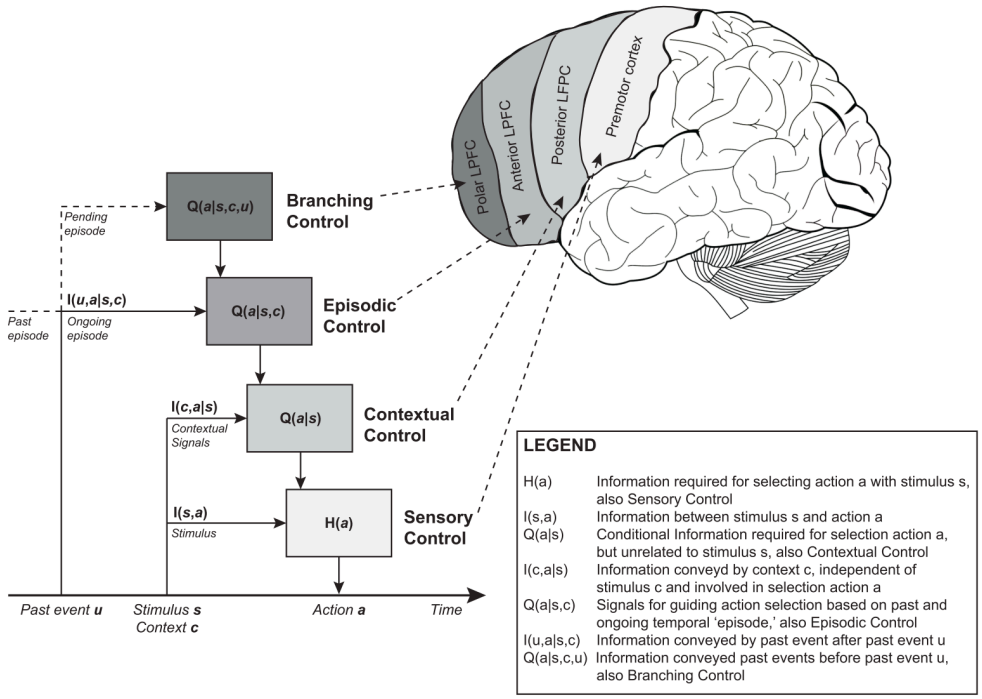


Figure 3.3: **The information cascade model by Koechlin et al**
The model posits that cognitive control operates according to three nested levels of control processes (branching, episodic, contextual), which are implemented as a cascade from anterior to posterior prefrontal regions.
Figure taken from [112]

model accounts for the effects of ACC damage observed on rodent behavior, and is compatible with the electrophysiological evidence of reward prediction errors, and neuroimaging data from effort and control studies, nevertheless, it does not account for the surprise and error signals observed in the ACC.

On the other hand, in the model proposed by Frank and Badre [77], the authors take the view that cortico-striatal-thalamic (loops between the PFC and BG) circuits are central in supporting the network interactions required for hierarchical control. They propose that these circuits are organized hierarchically, and in each circuit, the basal ganglia gate frontal actions, with some units gating input into the PFC and others gating outputs to facilitate response selection. This functionality of a nested output gating, with higher order information represented and maintained in rostral regions, which conditionalize attentional selection in more caudal regions, thus allows this hierarchical version of the PBWM model to learn conditional if-then problems.

More recently, Eckstein and Collins [68] have proposed a hierarchical RL model that consists of learning distinct selective models composed of stimulus action associations, as well as learning 'selective models of selective models' composed of associations between contextual cues and selective models, to arbitrate between the subordinate selective models and drive adaptive behavior across changing contexts. Collins and Frank [53] present a model that combines selective, predictive, and contextual models within task sets. This model implies that the selective model learns through model free RL and selects actions within the actor task set, while predictive models learn to predict action outcomes from selected responses to stimuli. Contextual models learn environmental cues to forecast proactively the reliability of predictive models. The reliability of task sets is computed using first order probabilistic inferences about the reliability of predictive models, both proactively and reactively based on actual outcomes. As long as the actor task set is judged reliable, it serves as a guide for behavior. In parallel, several other task sets are monitored and if one of them is regarded as more reliable, that task set is chosen to serve as the actor. If none of the monitored task sets are deemed reliable, a new task set is created by combining the task sets stored in long term memory.

This notion of reliability is not new. Doya, Samejima, Katagiri, and Kawato [67] proposed a multiple model based RL model that conceptualizes how several internal models within the actor task set might be weighted to drive behavior concurrently. Their model is made up of many pairs of selective and predictive models, called controllers. Predictive models are utilised to determine the reliability (called responsibility signal by the authors) of each controller, which is then used to (i) identify the weights or relative contribution of each selective model to action selection and (ii) adjust the updating of internal models across all controllers in re-

sponse to action outcomes. The model assumes that all controllers are identical in nature, although it might be modified to include controllers that operate at different time scales.

The *Hierarchical Error Representation* (HER) model by Alexander and Brown [3] (a hierarchical extension of the PRO model), explains cognitive control in terms of the interaction between the dlPFC (dorsolateral prefrontal cortex) and the mPFC (medial part of the PFC). The dlPFC learns to maintain representations of stimuli that reliably co-occur with outcome prediction error and these error representations are used by the mPFC to refine predictions about the likely outcome of actions. The error is broadcasted through the PFC in a bottom up manner, and modulated predictions from top-down facilitate selection of an appropriate response. Owing to its recursive architecture, this model can elaborate hierarchical rules on the basis of learning by weight updating, both to select pertinent stimuli and to map a representation inspired by the principles of predictive coding. In addition to its hierarchical structure, the model proposes to decompose the functioning of the PFC between, on the one hand, the prediction of the outcome and the monitoring of the error of prediction, and on the other hand, the elaboration of contextual (and possibly hierarchical) rules to compensate for errors. This distribution of functions has also been reported between respectively the medial and lateral parts of the PFC, as discussed in the previous chapter.

A computational model proposed by Khamassi, Lallée, Enel, Procyk, and Dominey [107] simulates cellular activity in both ACC and lPFC, on the basis that the feedback related signals in ACC (conflict or surprise) modulate the exploitation-exploration tradeoff in lPFC during decision making. Another model proposed by them focuses on the 'meta control', they provide a theory for how the interactions between the lateral and medial PFC play a role in the tuning of the meta learning parameters of the model, hence providing a solution to the problem of when to abandon an ongoing plan in favor of an alternate one.

Models of cognitive control that focus on ACC need to be integrated into more comprehensive accounts that explain how the lateral and medial parts of the PFC interact with one another and other brain regions. This thesis takes some of these models as the launch pad to investigate questions that existing models fail to account for.

Over the last two decades, many neuroscientists have become increasingly convinced that the brain works along the lines of bayesian logic. The idea being that the brain represents sensory information probabilistically, in the form of probability distributions. The existence of *something like* Bayesian predictions taking place within the nervous system to explain perception seems certain. For the moment,

the theoritical generalization of this assumption to explain the whole of the brain remains speculative. [103]

3.4 Modeling uncertainty

One missing aspect in the aforementioned models are the methods by which organisms notice environmental change and start the process of either switching between, or learning whole new behavioral strategies. Unexpected events and variable contexts also activate behavioral adaptation mechanisms in animals with highly developed nervous systems, including adjustments in foraging, action planning, motivational drive, valuation strategy, and the rate of picking up new information, such as rules and higher order statistical models of the environment. Future rewards are uncertain for a variety of reasons. Some rewards are risky or essentially probabilistic, which makes them uncertain (e.g., coin toss).

Such a process should be capable of adjusting to changes in the environment's statistical parameters (variability, outcome probability, and outliers), as well as its contingency structure (changes in the state space, its transition properties, and the introduction of new events). In typical RL algorithms, agents start with either a model of the world (model based algorithms) or just a collection of results seen in different states (model free algorithms), and they adjust the values assigned to the states and actions by a percentage proportional to discrepancy between their estimated and experienced values, the reward prediction error (RPE). Under very general assumptions, such algorithms are theoretically capable of locating at least a local optimum for the decision policy. Naive implementations may still need thousands of observations to converge on stable behavior due to the incremental udate process. However, such RL agents may thus find themselves perpetually playing catch up in an environment that rapidly alternates between numerous fixes, unique reward structures, unable to do more than gradually adapt in reaction to abrupt transitions. The ability of many species, including humans, to successfully adopt a large array of behavioral strategies, each of which may be independently altered and deployed with little switching costs, is obviously misrepresented by this situation. Consequently, despite the fact that many traditional theories of conditioning propose surprise (formalized as the absolute value of RPE) or similar violations of expectation as a way to dynamically adjust reward rates and promote new learning, such models are still predicated on the notion of a single policy that is subject to gradual updates. For instance, the performance of RL learning rules can be enhanced by adding extra meta learning parameters based on estimates of uncertainty, particularly since the values of its actions are changing. If agents adjust

their learning rates to reflect the environment, they can learn considerably more quickly. Contrast this with the difficulty of change detection, where it is necessary to distinguish between predicted variance in results and a real change in the environment's underlying structure [59]. In these situations, agents may perform better by learning a wide variety of behavioral strategies (or a meta approach with a limited number of fast adjustable parameters) and having the option of switching between them when a significant enough shift in the environment is recognized. Such a model operates similarly to more advanced models of conditioning that use bayesian mechanisms to dynamically alter learning rates, and it agrees well with the observed quickness of behavioral adjustment in the face of abrupt changes in the reward structure of the environment. In this case, outcomes would be monitored to assess whether the environment has altered significantly enough to call for a change in the existing strategy (and learning rates within that strategy). The change detection system, which would combine bayesian inference and a variety of innate or derived models of the world, would then function as a subprocess within reinforcement learning.

When uncertainty is high, it may be essential to extend the learning time scale so that full sequences of behavioraly important events are taken into account. This may require episodic memory and facilitate processes like transfer and one shot learning. Agents may place a special emphasis on "learning how to learn", maybe in part by synaptic adjustments that influence how brain circuits react to unexpected occurrences, or by altering their behavioral approach, such as, for example, facilitating various attentional modes. There have been suggested algorithmic approximations that depend on more clearly identifying a change point in latent states. Nassar, Wilson, Heasly, and Gold [144] condensed a Bayesian change detection model to a delta rule model, which modifies the influence of new results in accordance with the uncertainty and likelihood of change points. To address learning under volatility, there are additional Kalman filter inspired models that monitor both the estimates state of the system and the variance of these estimates.

Model	Model Type	Effects
Conflict Monitoring [27], [214]	Connectionist	Conflict, errors
Error likelihood [32]	Rate-coded neurons	Conflict, errors
Motor control filter [95], [94]	Reinforcement learning	errors, prediction, reward prediction error

Model	Model Type	Effects
Volatility [19]	Bayesian	Volatility
Choice Difficulty [26], [189]	Connectionist	Choice difficulty in decision making
Adaptive Effort allocation [202]	Reinforcement learning	Physical and cognitive effort and cost-benefit trade off
Expected value of control [187]	Conceptual	cost benefit trade off in decision making
Sychronization by oscillations [201]	Rate-coded neurons	Cognitive control by theta oscillations
PRO [4]	Rate-coded neurons, reinforcement learning	prediction and prediction error, conflict, error, pain
RVPM [190]	Rate-coded neurons, reinforcement learning	reward prediction and prediction error, conflict, error, volatility
HRL-ACC [96]	Reinforcement learning	effort, task switching, hierarchical behaviors
RNN-ACC [186]	Connectionist	distributed coding of extended action sequences, conflict, prediction errors
HER [3]	predictive coding, multi-dimensional error signals	working memory(dlPFC), hierarchical learning
ACC-LPFC [107]	Rate-coded neurons, reinforcement learning	reward prediction error, salience, exploration-exploitation trade off

Model	Model Type	Effects
PROBE [51]	bayesian	task-set creation and switching
Augment, Hybrid Augment [135]	Connectionist, reinforcement learning with attention gated memory	working memory, contextual learning
Rule set learning and selection [75]	Connectionist	rule set learning and selection, rule transfer
Temporal chunking [29]	RNN	task set learning and retrieval
WorkMATe [120]	Connectionist	cognitive control over working memory

Table 3.1: Summary of computational models, computational framework and effects on cognitive control addressed by them

3.5 Discussion

This chapter exemplifies some of the principal computational approaches to cognitive control in current research, including those seeking neuroscience applications of established formalisms from the machine learning literature, those that detail biologically inspired neural network models of interacting control mechanisms, and those deriving rational models of decision making and control from optimality constraints. Computational approaches can be used in both a reductionist and emergent manner : deconstructing the homunculus' mysterious intelligence into hopefully more understandable "dumb" neural subcomponents, while also demonstrating how complex control functions emerge from the dynamic interactions among these multiple sub components of cognitive control. A key aspect of this approach when applied to the domain of cognitive control is that these models view control as arising from the interactions of multiple relatively simple elements (e.g„ neurons or neural assemblies that perform local processes within a single brain system or unit). Thus, these models emphasize how cognitive control functions emerge from a network of brain regions activated interactively and in

parallel, rather than the more historical approach of localizing cognitive function to a single brain region.

Nevertheless, the models reviewed in this Chapter either explain the mechanisms of just one of multiple aspects of cognitive control (as in the case of working memory and monitoring (ACC) models) while ignoring their interactions with other computations involved in control. Or, in other cases (top-down or hierarchical models), they explain and implement more global constructs agnostically (task sets or hierarchical organization), while being specific only about a few of them (hierarchical loops of gating). In each of these models, we have found (i) one or more aspects of control (as detailed in Chapter 1, namely working memory, selective attention, hierarchical organization, and performance monitoring) missing or (ii) built on simplifying assumptions and partial explanations of behavioral data, or in some cases (iii) implementations that do not yet have a grounding mechanistic link in literature (backpropogation, bayesian models)

A key theme of the theory of 'method of minimal anatomies' is that a theory of an entire brain (or in the case of this thesis, Cognitive Control) cannot be derived in one step. Rather, this is done incrementally, in stages. A 'minimal model' is one for which, if any of the model's mechanisms is removed, then the surviving model can no longer explain a key set of previously explained data. In Chapters 4 and 5, taking inspiration from this theory, and from our conceptualization of Cognitive Control as a gradient, we implement and analyze computational models, increasingly appending them with the key ingredients needed for Cognitive Control.

In the first three chapters, we have walked through the current understanding of Cognitive Control through the lens of three fields : Cognitive Science (Psychology), Neuroscience and Computer Science. This has highlighted that a global framework for Cognitive Control, and hence any computational model that seeks to explain it, and further make predictions about behavior must have the following constituents, as per Marr's three levels of analysis :

- At the *computational* level
  Why is the computation that the cognitive system must perform important ?
  to deal with the challenges of catastrophic forgetting, the stability-plasticity dilemma, rule switching

- At the *algorithmic* level
  What is the algorithm for the transformation from inputs to outputs ?
  through mechanisms like working memory, selective attention, task sets, contexts, and hierarchical representations

- At the *implementation* level
  How do neural systems perform and carry out the cognitive functions described ?
  through biologically plausible mechanisms (prediction errors), rooted in neuroscientific data about connections between different brain regions (lateral-medial and anterior-posterior gradient)

In the following two chapters, we build computational models, that incrementally are able to bring in the constituents mentioned above, and this increasing gradient explains the gradient of cognitive control explained in Chapter 1, as seen in the complexity of observable behavior across different species.

Part II

Computational Modeling

4 Cognitive Control over Multiple Memory and Learning Systems

In this Chapter, we start providing the building blocks for understanding adaptive behavior and cognitive control in terms of different learning strategies that link stimuli, actions and outcomes to guide behavior. The concepts of different memory systems (working and episodic), performance monitoring, context, and default behavior that were invoked in Chapter 1, and their need and involvement for cognitive control, are detailed here, in a step by step manner, through simple computational models. The first section lays the groundwork for illustrating the different learning systems an organism (agent) has at its disposal, and the following two sections provide addendums to these systems, different aspects such as default behavior and context, that can either help or interfere with learning.

In the first section, we look at how adaptability arises from specific interactions between multiple learning and memory systems. We use an actor critic model as the base, to illustrate the capacity and limits of such a model, through increasingly complex decision making tasks taken from the rodent literature. Basic rules that link stimulus to a response or action (what we call concrete rules) are straightforward enough to learn through valued guided decision making. Particularly, the actor critic architectures for reinforcement learning have long been proposed as models of dopamine like reinforcement learning mechanisms in the rat's basal ganglia (together with learning mechanisms based on other subcortical structures like the midbrain and amygdala). Thus, this framework provided a solid foundation for the purpose of our investigation. Very quickly though, as soon as an element of time is brought into the picture, one faces what is called the temporal and structural credit assignment problem in RL. We show how the basic architecture we started with, can then be appended with a very minimal abstraction of a "working memory" like state to get around this issue. By the nature of its implementation, and its capacity limited resources, this yet again remains insufficient for more temporally extended tasks or ones that require the maintenance of several contexts, at

which point, an approximation of a long term or "episodic" memory is needed. Finally, we highlight that even with these systems in place, the model has limitations on its ability to explain complex behavior, and thus requires some mechanism to detect abrupt changes in the environment.

In addition to the role of the multiple learning systems noted above, organisms are not "naive", but have certain pre-learned "innate" behaviors, due to their unique social and ecological landscapes. Taking this into account, the computational processing challenge for an organism is not to optimize learning in a fixed evironment, but rather the challenge of lifelong, online learning, that must continuously cope with already learned behavior. In the second section, we highlight the influence of "default" behaviors in learning simple stimulus-response associations in a seemingly complex experimental setting. As explained in the conceptual overview, at its simplest, cognitive control is the ability to override default or dominant behaviors. However, most computational models in literature begin learning with either a blank slate, or with randomly initialised action values. Using the same actor-critic architecture as in section 1, we show the influence of innate behaviors in learning even basic concrete rules, or SR associations, given a sufficiently complex task setting. We use the findings and learning curves from an experimental task to validate the results of our model, thus highlighting an organism's need to "cope" with innate behaviors to demonstrate adaptability. We discuss the interpretations of the experimental result, and the different hypotheses about the neural mechanisms of inhibiting default behaviors. While competing accounts are proposed by different researchers, a unifying theme that emerges out of this study, is the need for a "context" representation for behaviors that do not cooperate but rather compete with the default behavior of an organism.

Thus, in the final section, we highlight the importance of context in learning and cognitive control. Even in situations where the context is overtly present, one needs to learn to identify the said stimulus as a "context". In cases where the context is covert rather than overt, this requires inferring it as a latent state. Using the same spatial alternation paradigm introduced in the first section, albeit in a continuous state space, we show that an agent can solve the alternation task without an explicit working memory, using a reservoir computing framework. Analysis of the model's internal activity reveals that the memory is encoded inside the dynamics of the network. However, such dynamic working memory remains inaccessible such as to bias the behavior of the agent into one of the two attractors (left or right). To do so, external cues are fed to the network such that the agent can follow arbitrary sequences, instructed by the cue. The model highlights that procedural learning and its internal representations can be dissociated, with the former being insuffi-

cient to allow for an explicit and fine grained manipulation required for cognitive control.

4.1 Deciphering the contributions of episodic and working memories in increasingly complex decision tasks

[1]

Learning and decision making are fundamental aspects of cognition and are closely linked. They have long been studied in animals through various levels of complexity in behavioral tasks. Reinforcement Learning (RL) provides a theoretical framework for modeling tasks in which agents interact with their environment and learn rules by receiving reward signals upon taking actions, and has an undeniable biological basis [124]. It is nonetheless constrained by the Markovian property, stating that the decision can be directly made from the present state, not consistent with known characteristics of animal behavior in real world situations or even in cognitive tasks.

This class of partially observable Markov decision problems (POMDPs) has been solved by extending the present state (representing the state of the environment) with internal representations [157] that might correspond to memories built from previous experiences. A basic version of this principle has been proposed in [215] and related to biological basis, with the distinction between a working memory (associated with the prefrontal cortex) [210], where a given cue can be kept present in memory for some time, even if it disappears from the experienced world, and an episodic memory (associated with the hippocampus), where a previous episode (series of steps) can be recalled by similarity from the present state and manipulated as a virtual state.

On the computational and experimental neuroscience sides, a biological neural network framework was proposed [150] to explain how working memory representations in the PFC may be updated and maintained. This concept of gating models was also used to study rule acquisition in rats [129], by comparing the ability of two RL algorithms to replicate rat behavior. The ability and limitations of such a model in a common human behavioral task has also been demonstrated [197]. More recently, a simplification of this model has been extended [105] to include a bias that better explains the performance of rats in a spatial alternation task.

[1] This article was presented at IJCNN, 2021 [56]

On the machine learning side, learning and exploiting these forms of memory have been adapted to RL [25], introducing complex representational and computational mechanisms that have also been compared in more details with human brain circuitry, thus introducing meta-RL [206] and episodic-RL [171] as new paradigms for addressing non Markovian problems. But this impressive level of performance is at the price of complex computations, requiring an often prohibitive training time (and correspondingly corpus size) and resulting in obscure computing phenomena, difficult to interpret and analyse in terms of functional contributions of the respective memory mechanisms.

What we propose here is to design a study where a basic RL agent is extended with a minimal version of working memory and of episodic memory, that can be trained quickly and without hyper-parameters, and to define tasks where the usefulness of each memory can be analysed in details. Particularly, what we want to understand and share with our experimental neuroscientific colleagues are the conditions where one or the other kind of memory are needed and where they are not sufficient and should be complemented with more complex mechanisms. In other words, this explanatory study is the premise for predictions to be confirmed in forthcoming experimental studies in neuroscience and for precise specification to be implemented in more powerful learning algorithms for machine learning.

4.1.1 Methods and Tasks

Computational model and architecture

Basic RL agent A tabular, actor critic temporal difference learning architecture with ϵ greedy exploration was used. The agent maintains a table of state values V, where $V(s)$ is the agent's current estimation of expected, temporally discounted reward that will follow state s. The agent also maintains a table of action values Q, where $Q(s, a)$ is the value of taking action a in state s. The *actor* part of the architecture follows a simple policy where the agent picks the action with the highest Q-Value, except with a probability $0 \leq \epsilon \leq 1$ the agent selects an action at random. In the *critic* part of the architecture, the TD error δ is computed when the agent takes an action a in state s and transitions to state s' after receiving a scalar reward r :

$$\delta = r + \gamma V(s') - V(s) \tag{4.1}$$

where $0 \leq \gamma \leq 1$ is a temporal discounting factor. The old state value and action value are then updated as :

$$V(s) \leftarrow V(s) + \alpha\delta \tag{4.2}$$

$$Q(s,a) \leftarrow Q(s,a) + \alpha\delta \tag{4.3}$$

where α is a learning rate parameter.

The agent acts on the state space $S = S_L$ where S_L are all the possible location states an agent can find itself in (numbered squares in a T-shape grid world), by taking motor actions $A = up, down, right, left$ which can result in a change of the location state.

RL AGENT WITH WORKING MEMORY : To include a working memory representation, the state space was augmented with an extra memory element. A factored state space representation was used : $S : S_L X S_{WM}$ with each tuple of the form (S_L, S_{WM})(grid world location, working memory contents) having its own set of action values. The total number of possible WM states is one more than the number of location states (one extra for empty memory at the start of the episode) i.e. $|S_{WM}| = |S_L| + 1$. The memory element is "hidden" as the agent can only access this state using the update action - which sets the working memory state to the current sensory location state, until it is overwritten when the next memory action is taken.

RL AGENT WITH EPISODIC MEMORY : We include an abstraction of the episodic memory system in the model, which is content addressable and temporally indexed. The model maintains a history of the agent's *n* most recently visited states, in order. After each action that changes the agent's location state, the previous state is added to the end of the list, and the oldest state in the list is removed (no state is removed for the first *n* steps). The model now has a 3 element tuple for state representation. The factored state space $S = S_L X S_{EP} X S_{WM}$. The episodic memory state S_{EP} can take either one state from the episodic memory list or an additional state representing "nothing recalled". To interact with this memory system, the agent can take two actions - "cue retrieval" to find the most recent instance of the agent's current state in the list (and set S_{EP} to that state) or "advance retrieval" to set S_{EP} to the state following its current state in the list. This kind of abstraction allows the model to retrieve a specific episode from its past and replay the memory from the retrieved point.

Tasks

Task A : Discrimination In the first, tactile discrimination task, the agent receives a sensory stimulus or a "cue" at the starting state. In this version of the task, it was a tactile cue about the surface, which could be rough on the right or left (represented by different states, as in Figure 4.1.B) and was indicative of the rewarding arm i.e. if the surface was rough on the right, the reward was placed at the end of the right arm while if it was rough on the left, the reward was placed at the end of the left arm. Thus, the agent's choice depended on learning this contextual or sensory rule. Another version of this task could be one where instead of a tactile stimulus, the agent could receive an auditory or odor stimulus in the starting state [31]. The important point is that this kind of sensory cue allows the rat or agent to form a distinct representation of the state

Task B : Alternation The second task is a spatial alternation task, in the environment as shown in Figure 4.1.A. This class of tasks is widely used to study hippocampal and working memory (PFC) functions [76]. The agent begins in the bottom of the central hallways (square marked with a black dot) and proceeds up to the choice point. On the first (sample) trial, the agent can either turn left or right and receives a positive reward at the end of the arm (squares marked highlighted in black). The agent then continues along the return arm and back to the starting point (where it is prevented from entering the side hallway by a barrier). Following the first trial, the agent receives a positive reward if it chooses the opposite direction as on the previous trial, otherwise it receives a negative reward.

Task 3 : Radial Maze In this task, there are three task conditions or contexts as represented in Figure 4.2 (left). In each task condition, the agent has the option of choosing only between 2 arms, with the rest of the arms blocked (the visual barriers being the contextual or sensory cue). For each of the contexts (A,B,C or D), the agent has to learn the alternating rule. In the trial phase, each context is presented once - A, B then C. The agent is free to go into any of the 2 open arms, and receives a reward at the end of the arm. Following the trial phase, the contexts are presented at random and the agent only receives a reward if it chooses the arm that it had not picked in the previous trial of the same context.

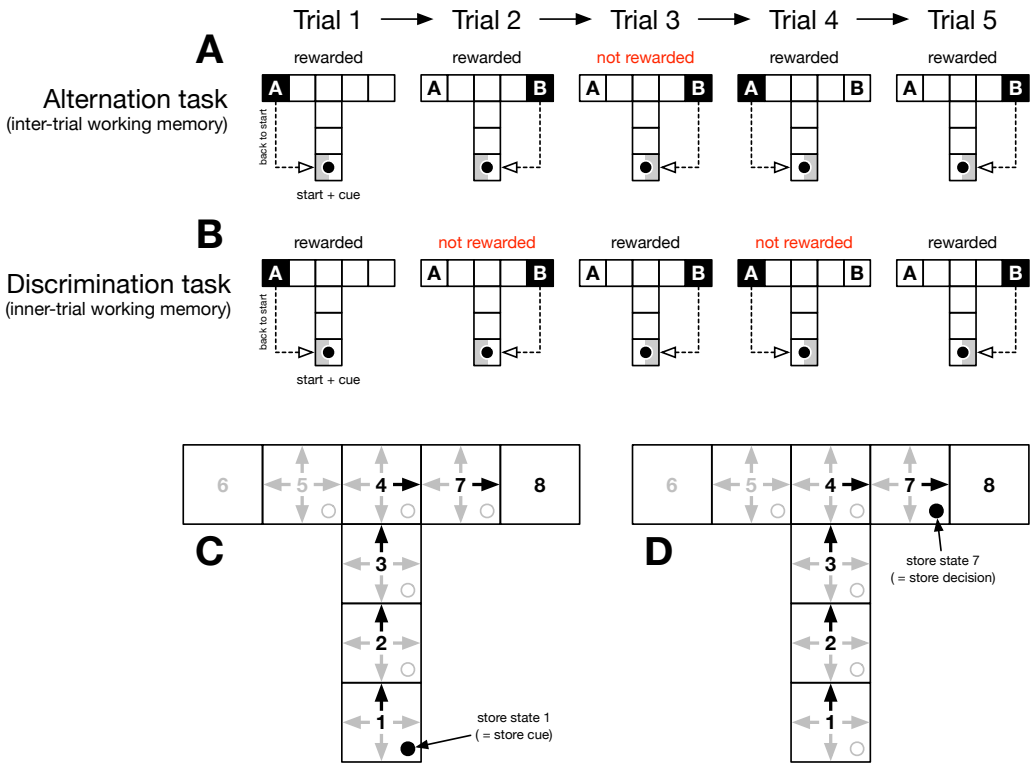


Figure 4.1: **A** The alternation task requires for an agent to alternate its choice between two different options (A & B). In the displayed setup, the environment is a classical T-maze where the agent starts from the bottom location in order to reach A or B. The first choice is free and the reward is obtained after each alternation between A and B. After each trial, the agent restarts from the initial location. On this example, the 5 trials can be written as ABBAB and only transitions AB and BA are rewarded. This task implies for the agent to remember its previous choice across trials. **B** The discrimination task requires for an agent to learn which cue (out of two) is associated with a reward. The agent has to choose between the two different options (A & B) depending on a cue that is presented at the entrance of the T-Maze. This task implies for the agent to remember the initial cue until it reached the corresponding target during a single trial. **C** One example for the discrimination task where the agent has learned to memorize the location at the entrance. **D** One example for the alternation task where the agent has learned to memorize the location after its choice has been made.

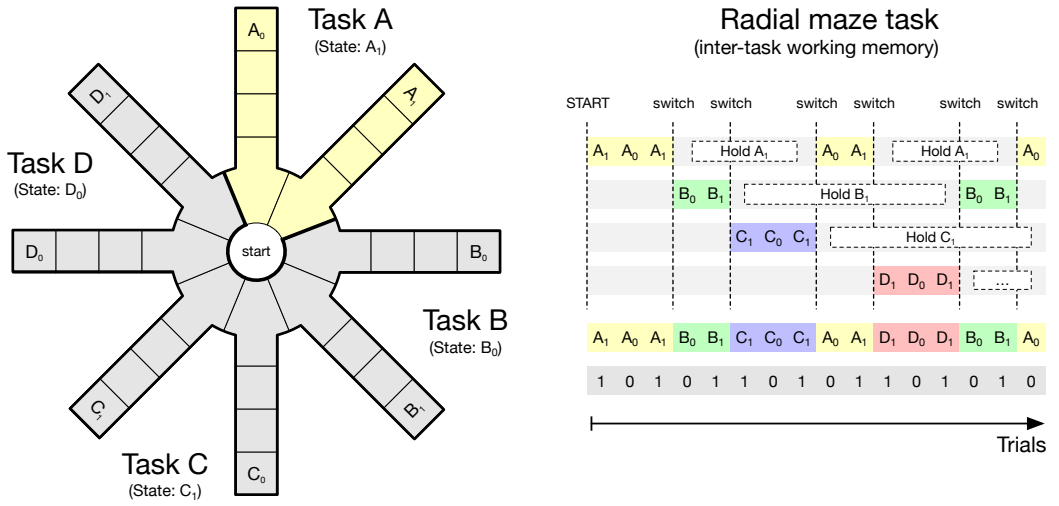


Figure 4.2: The radial maze task corresponds to a contextualized alternation task with four different contexts (A, B, C, D) as illustrated on the left part of the figure. At any time, only one context X is open such that the agent has only to choose between X_0 or X_1. The difficulty however is to maintain a memory for a given context when the context is changed. To be able to successfully solve this global task, the agent has thus to maintain simultaneously four different memories corresponding to the four different states of the context.

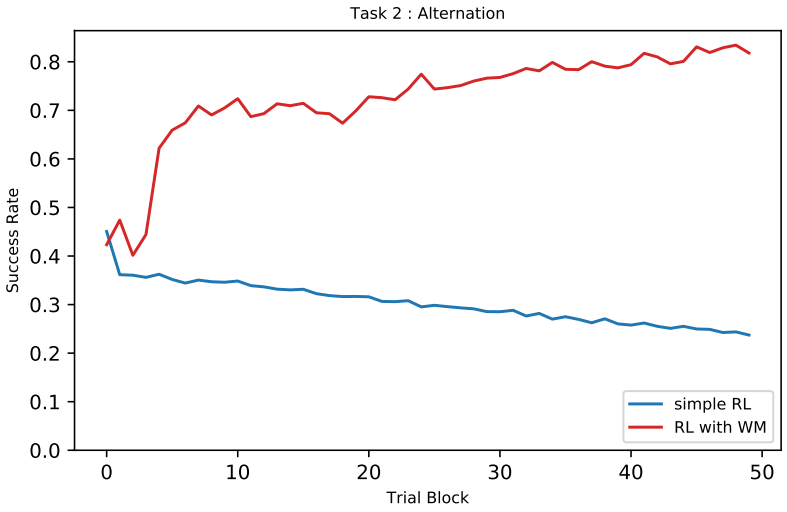
4.1.2 Results

Need for episodic memory

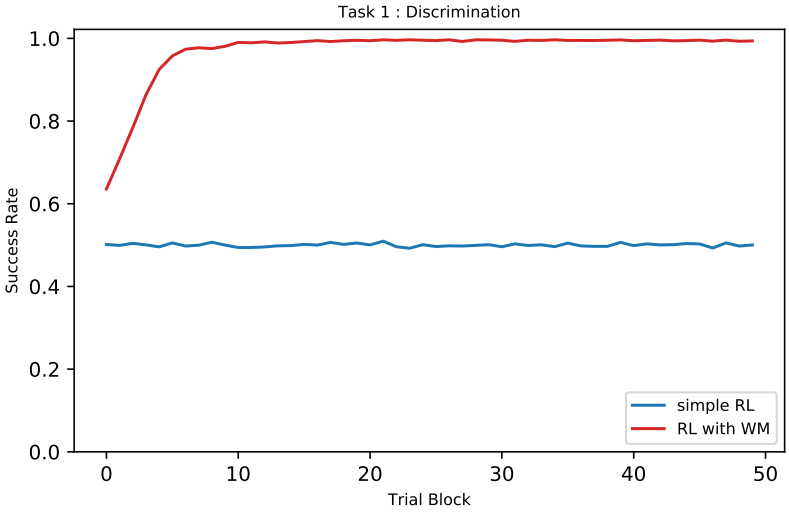
We first show, individually through the discrimination and alternation tasks, that the simple RL agent is unable to learn these tasks, remaining at chance performance for the discrimination task, and below chance for the alternation task. Performance in this task is plotted in Figure 4.3 (a) and (b). The RL agent with a working memory reaches an optimal performance by appropriately updating its internal memory at the right time. The agent learned different policies for each item that may be stored in the working memory. Accordingly, this can be interpreted as learning behaviors for a given context. In the discrimination task, the agent does this by buffering the sensory cue it received at the start state (or central arm), and maintaining it in memory until the choice point, thus disambiguating the trial type. In the alternation task, the agent had to buffer a location following the choice, and maintain this in memory until the choice point in the succeeding trial.

In the simulations, we could often observe that the agent could find the corresponding strategy and could learn to update and maintain the memory with the previous choice (for alternation) or the cue (for discrimination), whether left or right (accordingly, we call this strategy "remember both") (Figure 4.5 Right, Figure 4.6 Right). Interestingly enough, we could also observe sometimes the elaboration of another strategy, which is valid even if based on a side effect. In this "remember one" strategy, it is sufficient for the agent to remember only one choice (for alternation) or one cue (for discrimination) (Figure 4.5 Left, Figure 4.6 Left) and to simply label the other case by clearing the working memory. What is important at the end is to learn to associate the good action with a non ambiguous encoding of the state. The strategy adopted by the agent as a percentage over 100 runs is shown in Figure 4.4.

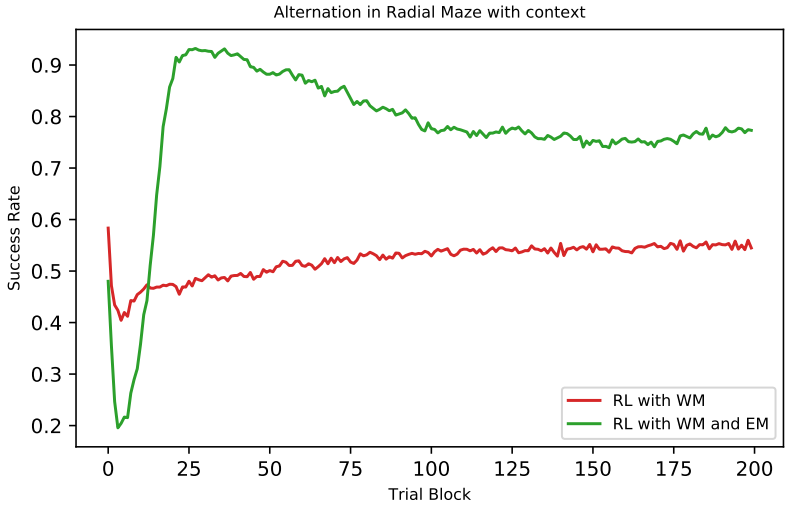
Next, we demonstrate the agent's behavior when the task is made more complex, and the agent has to learn the same behavioral rule of alternating between its choices, but under 3 different contexts. We use the radial maze environment and show that a simple augmentation of working memory is not enough to learn this task. While in a task with only two contexts, the RL agent with one working memory element may be able to perform the task at slightly above chance performance, increasing the number of tasks quickly deteriorates the learning. A separate memory mechanism, or the episodic memory is needed to learn the presented task, as shown in Figure 4.3 (c). To perform this task correctly, the agent depends on its episodic memory to retrieve the last presentation of the current context, advance one step in the memory to discover which direction it had previously chosen, and then choose the opposite direction. It may be possible to solve this task by increasing the number of memory elements, but this would exponentially scale the state space, making value learning increasingly difficult. Learning to use working memory is an implicit kind of learning, which evolves over time as knowing *what* and knowing *when* to update and maintain while making the use of episodic memory is an explicit learning of recall.



(a)



(b)



(c)

Figure 4.3: Performance of the agent in the **(a)** alternation task - a simple RL agents performs worse than chance as repeated visits to the same arm result in a negative reward **(b)** discrimination task **(c)** radial maze task. Each plot corresponds to the mean performance of 100 individual agents.

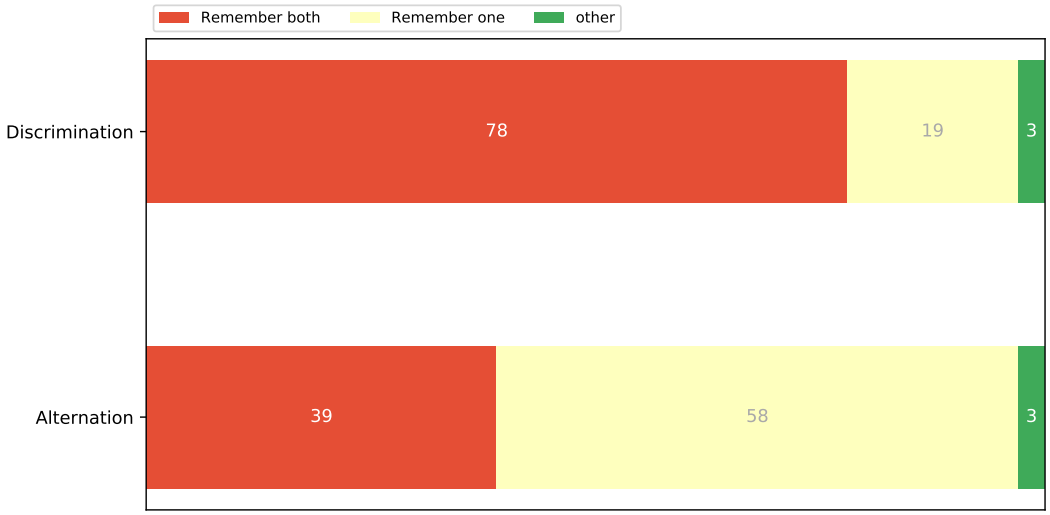


Figure 4.4: Percentage of strategy types learned in the Alternation and Discrimination tasks. The proportion of trials in the final block in which the previous choice (alternation) or the cue (discrimination) was in memory at the choice point was calculated. A threshold of 2/3. was used, for example, **(a)** the number of trials in which cue1 was in memory / number of trials with cue1 was greater than threshold, strategy was "remember cue1" (similarly for cue2) and **(b)** number of trials is which left was in memory / number of trials when previous choice was left was greater than threshold, strategy was "remember left" (similarly for right); if proportions for both were above threshold, the strategy was "remember both"; and if both were below threshold, it was "remember other".

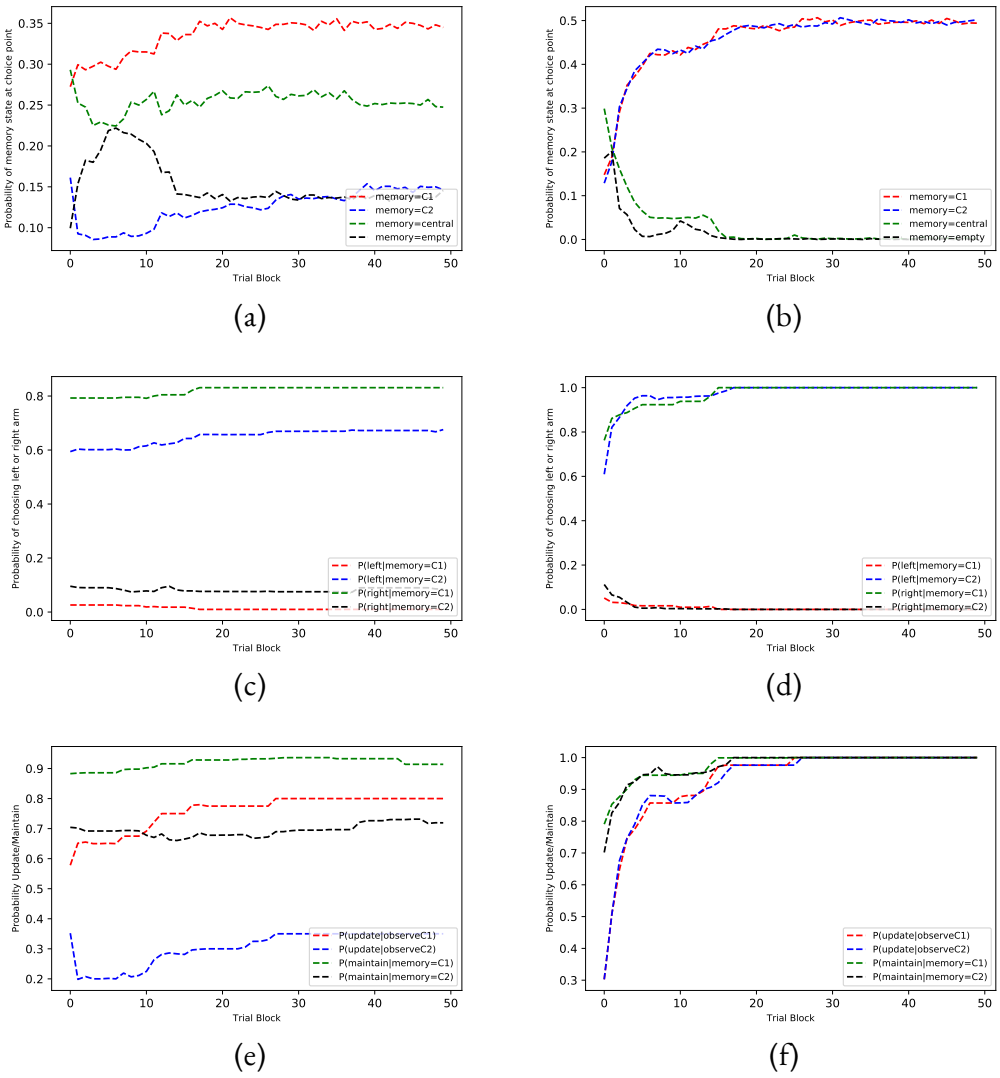


Figure 4.5: **Examples of "remember one"(left) and "remember both"(right) strategies learned by the model in the discrimination task. (a) and (b)** Probability over blocks of different possible memory contents (cue1/2, central arm or empty) at the choice point. Since on average, half of the trials begin with an observation of cue1 and the other with an observation of cue2, the maximum proportion of trials that either of these can be in memory is around 50% **(c) and (d)** Probability over blocks of choosing to turn right or left as a function of different possible memory contents. Probabilities are derived from Q-values at the end of each block. **(e) and (f)** Probability over blocks of updating or maintaining memory contents conditional on (1) observing the cue (at the start state) or (2) having it already in memory (before the choice point). Probabilities are derived from Q-values at the end of each block

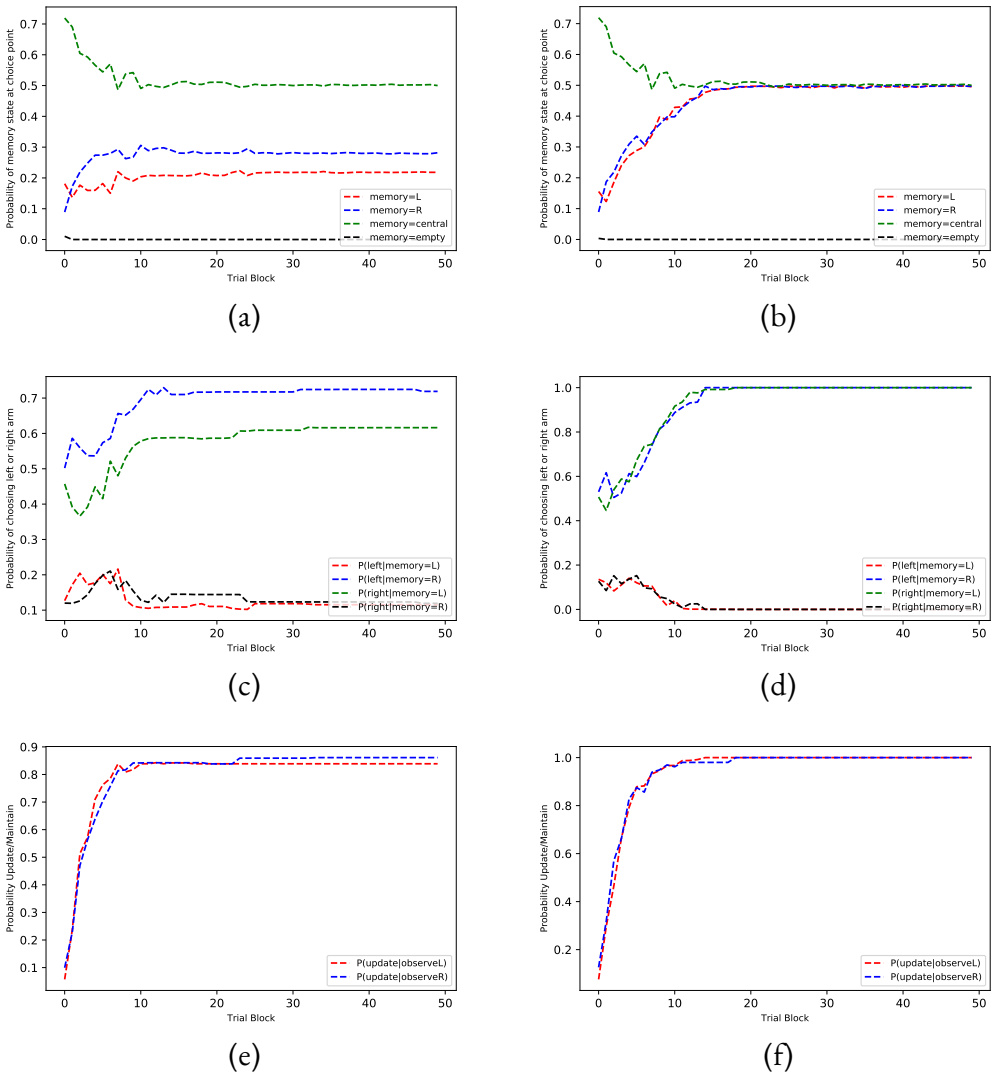


Figure 4.6: **Examples of "remember one"(left) and "remember both"(right) strategies learned by the model in the alternation task.** Probabilities plotted are the same as described in Figure 4.5, with observations of cue1 and cue2 replaced with observations of turning left or right in the previous trial

Need for performance monitoring / Transfer of learning

We analyzed the ability of the agent to adapt in a non-stationary environment. We tested if the agent is able to acquire the two distinct rules of task A (discrimination) and task B (alternation) by switching the underlying rule from one to the other, starting with task A, in the order ABAB. We show that after learning the first task,

the agent is able to learn the second task (alternation), but not to an optimal level of performance (as shown in Figure 4.7). This is due to contextual interference from the first task, in which the agent learned to 'pay attention' (i.e store in memory) to the presented cues. Nonetheless, due to the variation in the type of strategy used by the agent, it is able to reach a sub-optimal, but above chance level of performance. In the two rules we consider, the state spaces are only partially overlapping, and the difference in the policies is about *when* to update, thus the agent's performance doesn't drastically drop after the second switch. However in a situation where the rules are reversed [129] (for example, for the discrimination task - initially if cue1 rewards the right choice and cue2 rewards the left choice, a rule reversal would mean cue1 rewards the left choice and cue2 rewards the right choice), using the presented model would show such a sudden drop. This is because *(a)* the same set of Q-values are learned and updated continuously as the rules change, and *(b)* the model uses the same agent for learning motor and memory policies (as opposed to some multi-agent RL models). In any case, this model has the limitation of being unable to recall a previously learned rule. Hence, the best it can do is to identify a change in the environment, 'forget' its currently held policy, or adapt its exploitation/exploration, and relearn its action preferences.

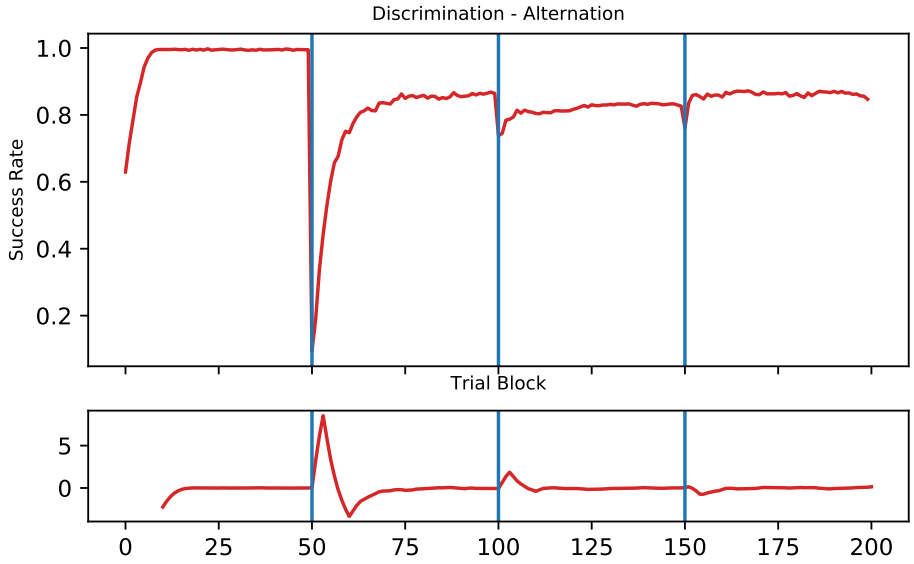


Figure 4.7: The agent learns task A (discrimination) for the first 50 episodes, then task B (alternation) for the next 50 episodes, followed by a switch to task A and then back to task B. The presentation of the cues interferes with the learning of the tas k. A simple method of performance monitoring to identify a rule change can be maintaining a running mean of the rewards obtained over the last n episodes

4.1.3 Discussion

The current study explored the ability and constraints of a RL model, augmented with working and episodic memories, to model rule learning under increasing levels of task complexity. Our focus for this work was to identify the limits of a minimal RL model, and increasingly complement it with biologically plausible mechanisms to explain learning behavior.

Even though the three tasks considered here are not really ecological, they are indeed experimented with rodents by our neuroscientific colleagues and could be used to replicate the present findings and explore possible predictions. Particularly, they allow us to illustrate the necessity of having multiple memory systems (working memory and episodic memory) in this context. For the simple discrimination task, which shares a number of similarities with a regular delay match to sample (DMTS) task, we explained that the Markovian Property forbids a basic RL agent to solve the task and an additional hidden state (working memory) is necessary to

reach the optimal behavior. The alternating task shares a lot of features with the discrimination task but the main difference concerning the working memory is not "what to store" but "when to store". We also showed that it is necessary for the model to store the state after the decision has been taken, that is, when the agent is close to one of the two end states. We also noticed that the model can efficiently learn to update and maintain working memory representations which can be intra-trial, inter-trial and inter-task.

Finally, the radial maze setup implies (in our implementation) a different kind of memory (episodic memory) such as to be able to "upload" the right context to the working memory. An alternative implementation could have been to have four different working memories, one for each context. But even in such case, the core "routing" problem remains the same: depending on the context, the agent needs to store the memory at the right place.

If we now go back to behavior, we think that the radial maze task is actually quite representative of our day to day behavior. An agent already engaged in a given task may "pause" it in favor of another task provided it perceives some cue indicating that a new task can be advanced or completed. This kind of behavior requires in fact the temporal organization of behavior. Furthermore, even though the task is already quite complex, we nonetheless keep it simple by explicitly cuing the agent with the unambiguous identification of the context, contrarily to tasks such as the Wisconsin Card Sorting Task (WCST) [194] that require an effective cognitive effort to monitor performance to decide if the context has changed or not. It thus comes as no surprise that our human brains possess dedicated areas in the frontal cortex to monitor, select and instantiate such high level rules.

Much work has been done on dealing with non-stationary environments, in model-free algorithms [55] in multi-arm bandits, or model-based RL frameworks such as that proposed by [67] but the neural underpinnings for these are not yet clear. On the other hand, models for rule-learning and rule-switching have been proposed that rely on bayesian inference [130] [51], and while they provide a unifying theory for these concepts, their complexity makes them infeasible for guiding and testing fundamental hypothesis in experiments. Coming back to biological inspiration and building upon working memory and episodic memory a biologically inspired cognitive control [69] [188] could be an interesting way to define a more flexible decision making agent, able to quickly adapt in an uncertain and changing world, as we will study in the near future.

4.2 Cognitive Control over Default Behaviors

[2]

The case of spatial alternation in mice, discussed above, merits special examination because it falls under what Watkins [207] refers to as "innate knowledge" (and what we have described as default behavior). It is a dependable spontaneous behavioral tendency that has likely been preserved through evolution due to some reproductive advantage it delivers to the organism. "What kinds of innate knowledge do animals have, and how does this innate knowledge contribute to learning?" is the central query posed by Watkins, in the context of learning and the assessment of learning rates. Furthermore, it is equally crucial to position this idea of "innate knowledge" in relation to the behavioral affordances offered by a certain context. It is highly relevant also in relation to cognitive control, since our understanding and explanation of cognitive control, first and foremost, is based on the ability of an organism to overcome default behaviors when they prove undesirable in a given environment. For instance, mice tend to spatially alternate in a T or Y maze, even when the behavior is not positively reinforced, leading to the conclusion that this particular state action policy can manifest itself even in the absence of a clear environmental reinforcer. Recent research has even demonstrated that mice spatially alternate in a T-maze even after they have already developed a preference for a reward that is only present in one of the arms [89].

Through reinforcement learning, we get formalisms for accounting for both the creation and revision of policies. In both instances, it is believed that this is the result of the agent assessing results (which may be fixed or dynamic) from actions taken while in a specific state, associatively storing these state-action-outcome evaluations, and using them to guide subsequent action while in the same (or similar) state. As a result, the term "state-action" policy enables us to group together under a single abstract idea all cognitive material that is known to control how an organism (or agent) chooses to act in a given situation (where beliefs, guidelines, attitudes, correlations between stimuli and responses etc comprise cognitive content).

According to the animal exploration theory, the principle cognitive drive underpinning exploratory behavior, is global information gain, over foraging [100]. Through this interpretation, it seems more plausible to explain the spontaneous

[2] This work was done in collaboration with Christopher Stevens (on the experimental side) during his PhD in Neuroscience, in the Inserm Magendie research unit, on the Bordeaux NeuroCampus. The experimental results have been published in [192]

spatial alternation behavior of mice as the need to explore. In fact then, it is the physical constraints of the maze (T or Y) environment settings that allow for this exploration to be interpreted as what researchers identify as "spontaneous spatial alternation". In terms of reinforcement learning, spatial alternation can be simply viewed as the most effective or ideal strategy for navigating the T or Y maze. On the other hand, since the environmental conditions of the tactile discrimination task explicitly reward non-exploratory behavior, this learning can be interpreted as requiring a context-dependent adjustment of the inherent exploratory policy rather than as the initial creation of a novel policy.

This is specifically what we investigate in this section, using the experimental framework of Stevens [192], and the modeling approach introduced in the first section. We show how the established default behavior of alternating (alternatively interpreted as foraging, curiosity, or information seeking), observed in rodents, interferes with the learning and subsequent expression of a tactile rule. While on one hand, this is a critique of most existing models, on the other hand, it provides an illustration of how a simple model can explain some of the behavioral results observed in more "ecological" settings.

4.2.1 Methods and Task

Computational Model

The same actor critic model, as described in the previous section was used, with an internal state or 'working memory' representation. The agent maintains a table of state values V, and a table of action values Q, as described.

For including the effects of the 'default behavior', or the spontaneous alternation, or the curiosity drive, or exploration, we used an innately learned policy $H(s, a)$, which was sensitive only to the history of selected actions, and not to the outcomes of those actions. To account for the preference of the agent to alternate in the maze, we added a bias to the propensities of the action values in $H(s, a)$:

$$H(s, a) = bias * H(s, a) \tag{4.4}$$

The bias function reduces the probability of selecting that action (or choice) which was made in the previous trial in the following manner :

$$bias(a) = \begin{cases} 1 & \text{if } s \neq s_c \text{ or } a_{trial-1} \neq a_i \\ 0.7 & \text{if } s = s_c \text{ and } a_{trial-1} = a_i \end{cases} \tag{4.5}$$

Finally, an action is selected by computing the weighted sum of action values from the learned and innate policies :

$$D(s,a) = w * H(s,a) + (1-w) * Q(s,a) \tag{4.6}$$

The weight w that governs the relative influence of each policy on the final action chosen, at each trial, remained fixed for the whole duration of the simulation.

Task

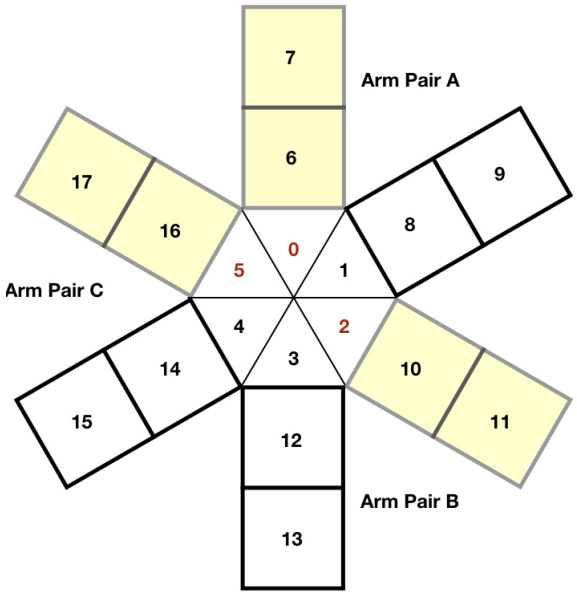


Figure 4.8: The radial maze task corresponds to 3 pairs of arms, of which only one within each pair has a surface associated with a reward (arms marked in yellow). The maze is modeled as discrete states, with the beginning state always in the central platform (0-5), as illustrated. The agent finds itself in a different starting state, depending on the relative presentation of the surfaces (state 0, in pair A, when the rewarding surface, S1, is on the left, or state 1, when the rewarding surface is on the right, and so on for other pairs

.

The radial maze task is modelled as in figure 4.8. There are three task contexts, in which one pair of arms (A, B, or C) is open at a time, and the agent has the option of choosing only between those 2 arms, with the rest of the arms blocked.

Each pair of arms has one rewarding surface (represented in yellow) S1 and one unrewarding surface S0. These surfaces are "fixed" in one session or episode i.e. if on session one, S1 was the left arm in pair A, on session 2 it might be on the right and so on for each pair). The agent only receives a reward at the terminal states of the arms. On the central platform, the agent can thus find itself in one of the six different states (0/1 when the context is A, 2/3 when the context is B, and 4/5 when the context is C). For each context, there is a binary state representation in the central (or starting) platform to represent S1 being on the right arm or left.

Thus, ultimately, the policy the agent must learn is to go right when in state 0, 2, or 4 and to go to the left when in state 1, 3, or 5.

4.2.2 Results

Experimental Task from Stevens et al

Acquisition and expression of a binary choice-based tactile discrimination of foraging rule

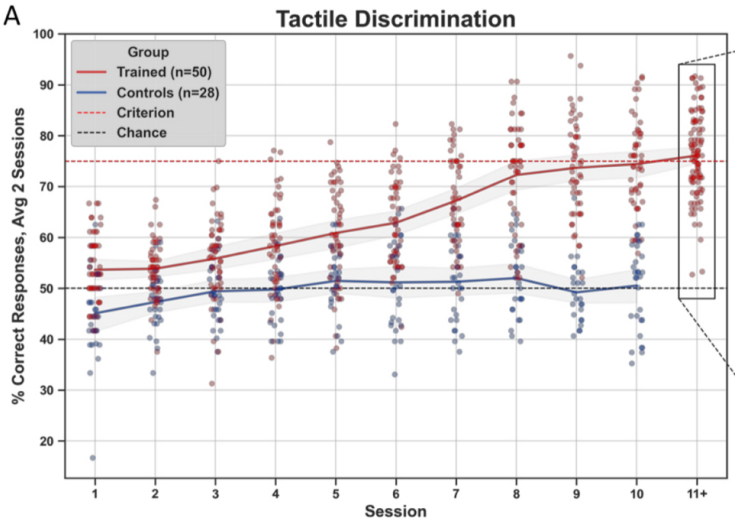
Mice were trained under a tactile discrimination based reward location association rule (R1) to choose, trial after trial, between the two contiguous radial maze arms of a sequence of arm-pairs, each arm of which was covered with one of two distinct surface types, one predictive of reward location (S1), one predictive of absence of reward (S0). The training was conducted in conditions of zero or almost zero visibility, i.e. without any extra-maze spatial cues, thereby constituting a classical stimulus-response (S-R) task where the stimulus in question was tactile. Across trials, the relative left and right position of S1 and S0 was counter balanced. Training consisted of either one or two sessions per day, with each session composed of between 16 and 36 trials. In early sessions, sequences were composed of a combination of repeated consecutive presentations of a same pair plus pseduo-randomized presentations of all available pairs. The aim of the repeated sequences was to explicitly lead the animal to inhibit its innate drive to alternate. As the animals approached criterion level across sessions, the arm pair presentation sequences became progressively pseudo-random. The final sessions of tactile discrimination

training were therefore fully pseudo-random trial sequences. Performance criterion was fixed as follows : animals had to attain either an average of at least 75% correct responses across two of the final pseudo-random sessions. In experiments with control groups, control animals were rewarded on every trial regardless of which surface, S1 or S0, they chose.

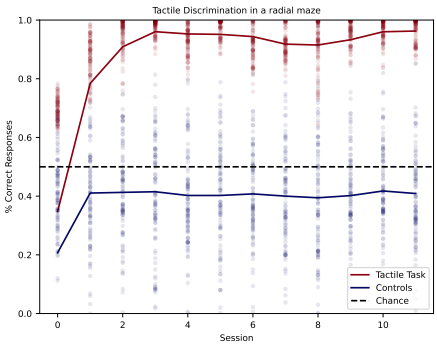
Computational Model

Simulation 1 :
The agent was trained to learn the simple tactile discrimination task as described in the previous section, from a blank slate ie with Q values initialized to 0. The performance is shown in figure 4.9 (b).

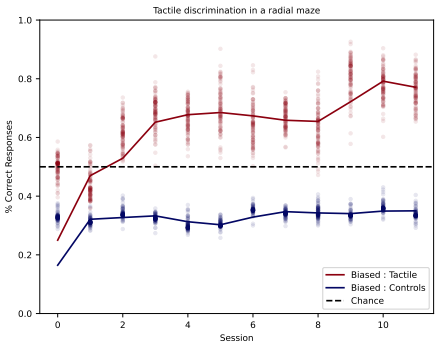
Simulation 2:
The agent was pre-trained as on the radial maze, as in the control condition (ie when there was a reward at the end of the arm, regardless of the surface chosen), so as to obtain a policy for the default behavior $H(s, a)$ (with equal preference for either surface), and this policy was then biased (as described in the previous section) such as to reduce the probability of repeating the previous choice, thus mimicking an alternation bias. This default behavior was assumed to be rigid, and when the model was subsequently trained on the tactile task, only the new policy values $Q(s, a)$ were updated with the TD error. The performance is shown in figure 4.9 (c).



(a)



(b)

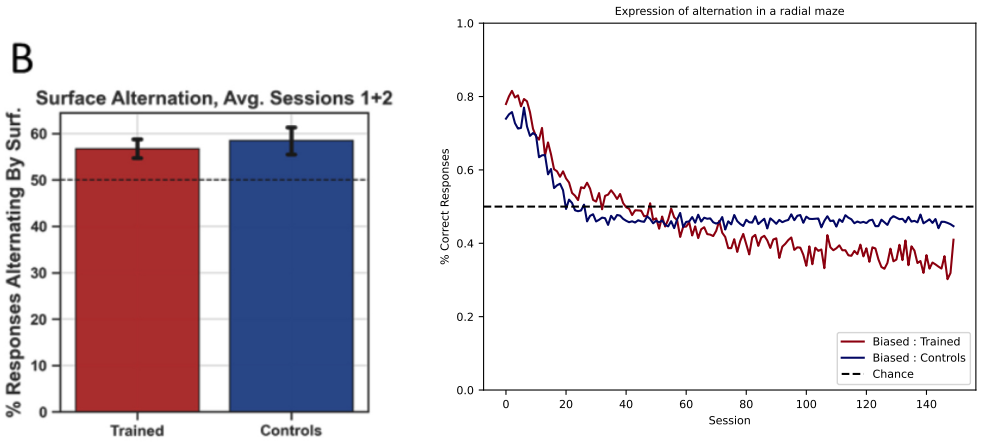


(c)

Figure 4.9: **(a)** % correct tactile discrimination responses over repeated training sessions, displayed as rolling averages over 2 sessions. R1 trained animals (n=50) represented in red, control animals (n=28) represented in blue. Controls were rewarded on all trials, regardless of surface chosen. Curves represent mean population score, dots represent individual performances. Figure taken from [192] **(b)** Performance curves of the model with no bias, in the control (in blue) and tactile trained (in red) simulation conditions. **(c)** Performance curves of the model with default behavior bias, in the control (in blue) and tactile trained (in red) simulation conditions. All the performance curves are averaged over 100 individual simulations of the agent. Each "session" is comprised of 5 blocks of 100 trials, and the mean of these 500 trials gives the % correct responses on the y axis. Dots represent individual simulations.

Figure 4.10 (b) shows the expression of surface alternation of the model with an added default behavior component, during learning. The model was trained on the tactile discrimination rule protocol. As in the results from the experiment (figure 4.10 (a)), initially the agent displays a greater than chance expression of the alternation behavior, in both the control and tactile trained simulation conditions, which eventually drops below chance over extensive training.

The results show that a model that is biased by a default behavior is able to better capture the behavioral results from the experimental study. Secondly, this model is able to capture the expression of alternation behavior observed in the initial training sessions.



(a) (b)

Figure 4.10: **(a)** Initial surface alternation behavior averages across first 2 sessions in both experimental and control populations. Figure taken from [192]. **(b)** Expression of the surface alternation behavior, during learning in the model biased with default behavior, in trained and control simulation conditions.

Simulation 3:

The biased agent was trained as in Simulation 2, but the bias was removed mid way through the training. The results (figure 4.11) show that the model is in fact learning by "coping with" the bias of the default behavior. When this bias is removed at the point when the model has learned the tactile task with a 80 % probability, the performance of the agent drops, proving that the tactile policy it has learned is

affected strongly by the default policy. This result qualitatively captures the experimental results which show that rule learning is disassociated from rule expression, with the latter strongly affected by innate or default behaviors. This effect is evidently not observed in a model that starts from scratch, with only one policy to learn, where the learning and expression of the rule occurs in parallel.

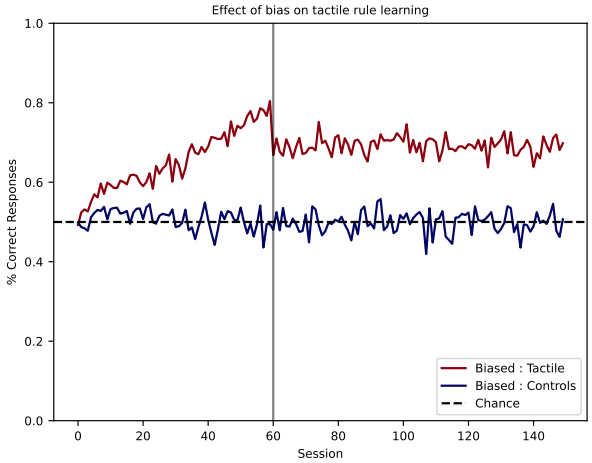


Figure 4.11: Expression of the tactile rule when the bias provided by the default behavior is removed after 60 episodes

4.2.3 Discussion

The study by Stevens et al, used in this section, was specifically designed to frame a behavioral training paradigm for rodents, which was purposefully antagonistic with respect to their spontaneous drive to explore. In so far as innate or "naive" animal behaviors can be viewed as evolutionarily maintained state action policies, of which "exploration" was the one rodents were expected to express in the particular task setting, the task can be regarded as a straightforward state action policy revision.

It is believed that innate behavioral traits of complex animals that have survived through evolution, like curiosity, reflect the kinds of cognitive responses most likely to provide an adaptive advantage in the variety of contexts in which a species has developed. When compared to innate, default behaviors, novel, learned behaviors can be either cooperative or competitive. These behaviors are acquired as

a result of contingent particularities of the environment the animal actually finds itself in. In light of this, cognitive plasticity or flexibility, or more generally cognitive control can be conceived as a trait of complex creatures, allowing them to specifically inhibit and go beyond innate behaviors, in acquiring and expressing new, learned behaviors such as to best exploit unpredictable environments.

A possible hypothesis which supports the results we show with our model, is the Heksor theory [212], which proposes the idea of a negotiated equilibrium between networks implementing newly acquired and old behaviors. The problem of "catastrophic interference", which is now prevalent in the research on artificial neural networks (ANN), illustrates the difficulties that a widely plastic system faces in learning a new behavior while simultaneously preserving existing ones (originally described by Grossberg [87] as the 'stability plasticity' dilemma). A second behavior may render an ANN incapable of producing the first behavior it had already learned. An adaptive behavior is based on a network of neurons and synapses that are constantly changing to improve performance and reduce errors, with the changes guided by feedback during the behavior and its outcomes.

While we don't make strong claims on the basis of this study, regarding the specific mechanisms that contribute to how shared neurons and synapses reach a negotiated equilibrium that allows for new and old behaviors to coexist, it is a possible explanation as to how various behavioral schemas, in the absence of explicit context, may coexist in the PFC.

The need to overcome default behaviors is obvious, and the role of inhibition has been highlighted, for context control learning. We have discussed in previous chapters, the critical need of inhibition for enabling controlled behavior : bad habits, unfamiliar situations, and dangerous environments in an animal's life often require that default behaviors be inhibited and more context appropriate actions performed. However, effective inhibitory control not only requires actually stopping unwanted actions, thoughts, or emotions - it also requires the efficient detection of those contexts that indicate the need for those forms of stopping. However, in the absence of an overt context, an animal must still cope with its innate tendencies. This is what we show in our model.

In the next section, we take this further and show how in the absence of explicit contexts, an agent can find itself unable to overcome its learned behavior, and provide an illustration for such contexts may then be developed internally.

4.3 From implicit learning to explicit representations

[3]

Suppose you want to study how an animal, when presented with two options A and B, can learn to alternatively choose A then B then A, etc. One typical lab setup to study such alternate decision task is the T-maze environment where the animal is confronted to a left or right turn and can be subsequently trained to display an alternate choice behavior. This task can be easily formalized using a block world as it is regularly done in the computational literature. Using such formalization, a simple solution is to negate (logically) a one bit memory each time the model reaches A or B such that, when located at the choice point, the model has only to read the value of this memory in order to decide to go to A or B. However, as simple as it is, this abstract formalization entails the elaboration of an explicit internal representation keeping track of the recent behavior, implemented in a working memory that can be updated when needed.

But then, what could be the alternative? Let us consider a slightly different setup where the T-Maze is transformed into a closed 8-Maze (see figure 4.12-Left). Suppose that you can only observe the white area when the animal is evolving along the arrowed line (both in observable and non-observable areas). From the observer point of view, in the central corridor, the animal is turning left one time out of two and turning right one time out of two. Said differently, the observer can infer an alternating behavior because of its partial view of the system. The question is: does the animal really implement an explicit alternate behavior or is it merely following a mildly complex dynamic path? This is not a rhetorical question because depending on your hypothesis, you may search for neural correlates that actually do not exist. Furthermore, if the animal is following such mildly complex dynamic path, does this mean that it has no explicit access to (not to say no consciousness of) its own alternating behavior?

This question is tightly linked to the distinction between implicit learning (generally presented as sub-symbolic, associative and statistics-based) and explicit learning (symbolic, declarative and rule-based). Implicit learning refers to the *non-conscious effects that prior information processing may exert on subsequent behavior* [47]. It is implemented in associative sensorimotor procedural learning and also in model-free reinforcement learning, with biological counterparts in the motor and premotor cortex and in the basal ganglia. Explicit learning is associated with conscious-

[3] This work was done with Naomi Chaix-Eichel, another PhD student in the Mnemosyne team and is available as a preprint [39]

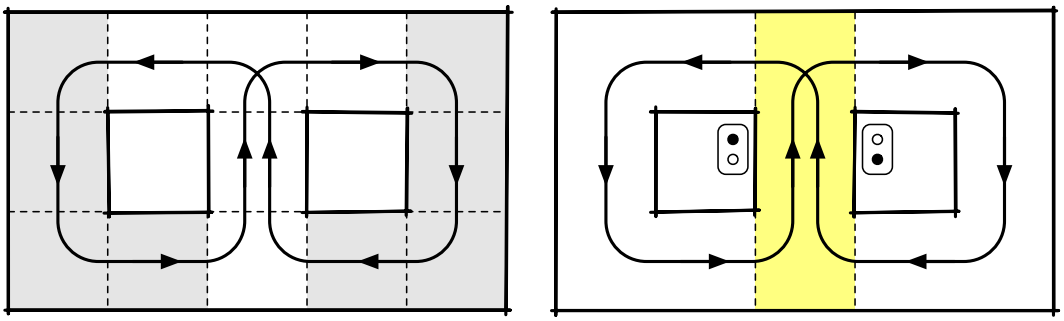


Figure 4.12: Left: An expanded view of a T-Maze. An observer can infer an alternating behavior because of her partial view (white area) of the system. Right: 8-maze with cues. A cue (left or right) is given only when the bot is present inside the yellow area.

ness or awareness, and to the idea of building explicit mental representations [48] that can be used for flexible behavior, involving the prefrontal cortex and the hippocampus. This is what is proposed in model-based reinforcement learning and in other symbolic approaches for planning and reasoning. These strategies of learning are not independent but their relations and interdependencies are not clear today. Explicit learning is often observed in the early stages of learning whereas implicit learning appears on the long run, which can be explained as a way to decrease the cognitive load. But there is also a body of evidence, for example in sequence learning [49] or artificial grammar learning studies [167], that suggests that explicit learning is not a mandatory early step and that improvements in task performance are not necessarily accompanied by the ability to express the acquired knowledge in an explicit way [62].

Coming back to the task mentioned above, it is consequently not clear if we can learn rules without awareness and then to what extent can such implicit learning be projected to performance in an unconscious way? Furthermore, without turning these implicit rules into an explicit mental representation, is it possible to manipulate the rules, which is a fundamental trademark of flexible adaptable control of behavior?

Using the reservoir computing framework generally considered as a way to implement implicit learning, we first propose that a simple alternation or sequence learning task can be solved without an explicit pre-encoded representation of memory. However, to then be able to generate a new sequence or manipulate the rule learnt, we explain that inserting explicit cues in the decision process is needed. In a second series of experiments, we provide a proof of concept still using the reservoir computing framework, for the hypothesis that the recurrent network forms con-

textual representations from implicitly acquired rules over time. We then show that these representations can be considered explicit and necessary to be able to manipulate behaviour in a flexible manner.

In order to provide preliminary interpretation of what is observed here, it is reminded that recurrent networks, particularly models using the reservoir computing framework, are a suitable candidate to model the prefrontal cortex[65], also characterized by local and recurrent connections. Given their inherent sensitivity to temporal structure, it also makes these networks adaptable for sequence learning. This approach has been used to model complex sensorimotor couplings [196] from the egocentric view of an agent (or animal) that is situated in its environment and can autonomously demonstrate reactive behaviour from its sensory space[9], as we also do in the first series of experiments, for learning sensorimotor couplings by demonstration, or imitation. In the second series of experiment, we propose that the prefrontal cortex is the place where explicit representations can be elaborated when flexible behaviors are required.

4.3.1 Methods and Task

The objective is the creation of a reservoir computing network of type Echo State Network (ESN) that controls the movement of a robot [8], [9], which has to solve a decision-making task (alternately going right and left at an intersection) in the maze presented in figure 4.12.

Model Architecture : Echo State Network

An ESN is a recurrent neural network (called reservoir) with randomly connected units, associated with an input layer and an output layer, in which only the output (also called readout) neurons are trained. The neurons have the following dynamics:

$$\mathbf{x}[n] = (1-\alpha)\mathbf{x}[n-1] + \alpha\tilde{\mathbf{x}}[n] \tag{4.7}$$

$$\tilde{\mathbf{x}}[n] = \tanh(W\mathbf{x}[n-1] + W_{in}[1;\mathbf{u}[n]]) \tag{4.8}$$

$$\mathbf{y}[n] = f(W_{out}[1;\tilde{\mathbf{x}}(n)]) \tag{4.9}$$

where $\mathbf{x}(n)$ is a vector of neurons activation, $\tilde{\mathbf{x}}(n)$ its update, $\mathbf{u}(n)$ and $\mathbf{y}(n)$ are respectively the input and the output vectors, all at time n. W, W_{in}, W_{out} are respectively the reservoir, the input and the output weight matrices. The notation $[.;.]$ stands for the concatenation of two vectors. α corresponds to the leak rate. $tanh$ corresponds to the hyperbolic tangent function and f to linear or piece-wise

linear function.

The values in W, W_{in}, W_{out} are initially randomly chosen. While W, W_{in} are kept fixed, the output weights W_{out} are the only ones plastic (red arrows in Figure 4.13). In this model, the output weights are learnt with the ridge regression method (also known as Tikhonov regularization):

$$W_{out} = Y^{target} X^T (XX^T + \beta I)^{-1} \tag{4.10}$$

where Y^{target} is the target signal to approximate, X is the concatenation of 1, the input and the neurons activation vectors: $[1; u(n); x(n)]$, β corresponds to the regularization coefficient and I the identity matrix.

Experiment 1 : Uncued sequence learning

The class of tasks called spatial alternation has been widely used to study hippocampal and working memory functions [76]. For the purpose of our investigation, we simulated a continuous version of the same task, wherein the agent needs to alternate its choice at a decision point, and after the decision, it is led back to the central corridor, in essence following an 8-shaped trace while moving (see figure 4.12-Left). This alternation task is widely believed to require a working memory such as to remember what was the previous choice in order to alternate it. Here we show that the ESN previously described is sufficient to learn the task without an explicit representation of the memory.

Tutor model In order to generate data for learning, we implemented a simple Braintenberg vehicle where the robot moves automatically with a constant speed and changes its orientation according to the values of its sensors. At each time step the sensors measure the distance to the walls and the bot turns such as to avoid the walls. At each timestep, the position of the bot is updated as follows:

$$\theta(n) = \theta(n-1) + 0.01 \sum_i \alpha_i s_i \tag{4.11}$$

$$p(n) = p(n-1) + 2 * (cos(\theta(n)) + sin(\theta(n))) \tag{4.12}$$

where $p(n)$ and $p(n+1)$ are the positions of the robot at time step n and $n+1$, $\theta(n)$ is the orientation of the robot, calculated as the weighted sum (α_i) of the values of the sensors s_i. The norm of the movement is kept constant and fixed at 2.

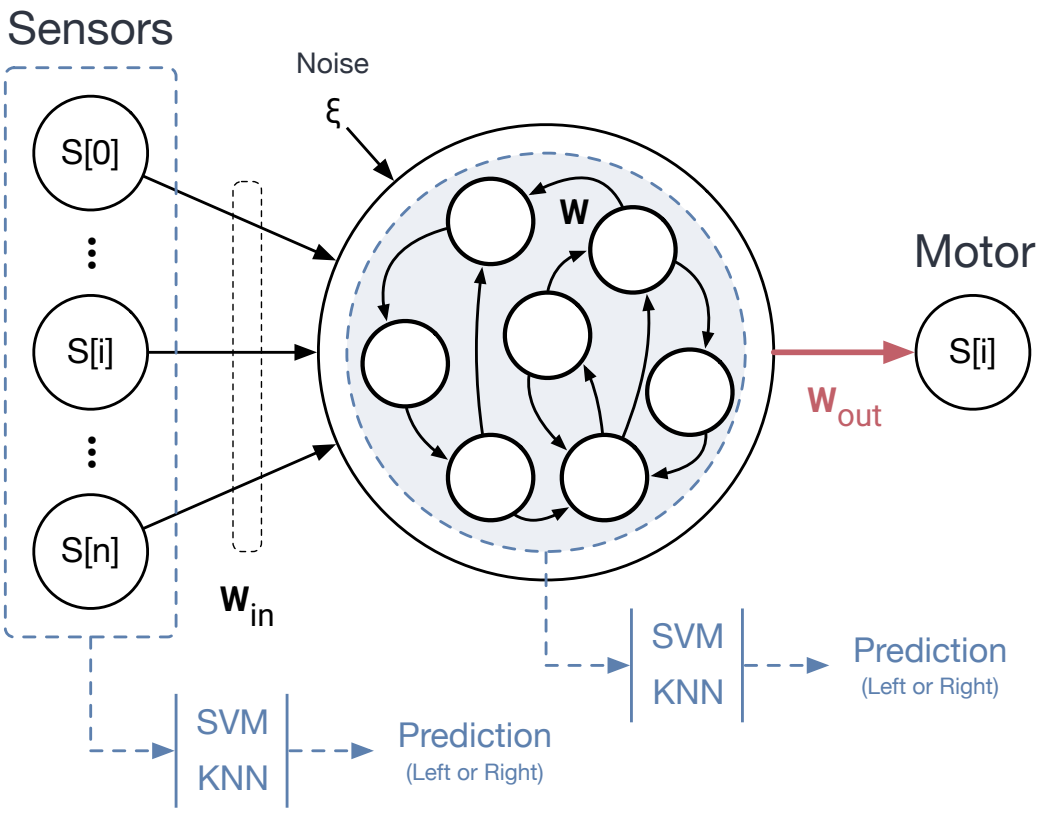


Figure 4.13: Model Architecture with 8 sensor inputs, and a motor output (orientation). The black arrows are fixed while the red arrows are plastic and are trained. The reservoir states are used as the input to a classifier which is trained to make a prediction about the decision (going left or right) of the robot. A left (L) and right (R) cue can be fed to the model depending on the experiment (see Methods).

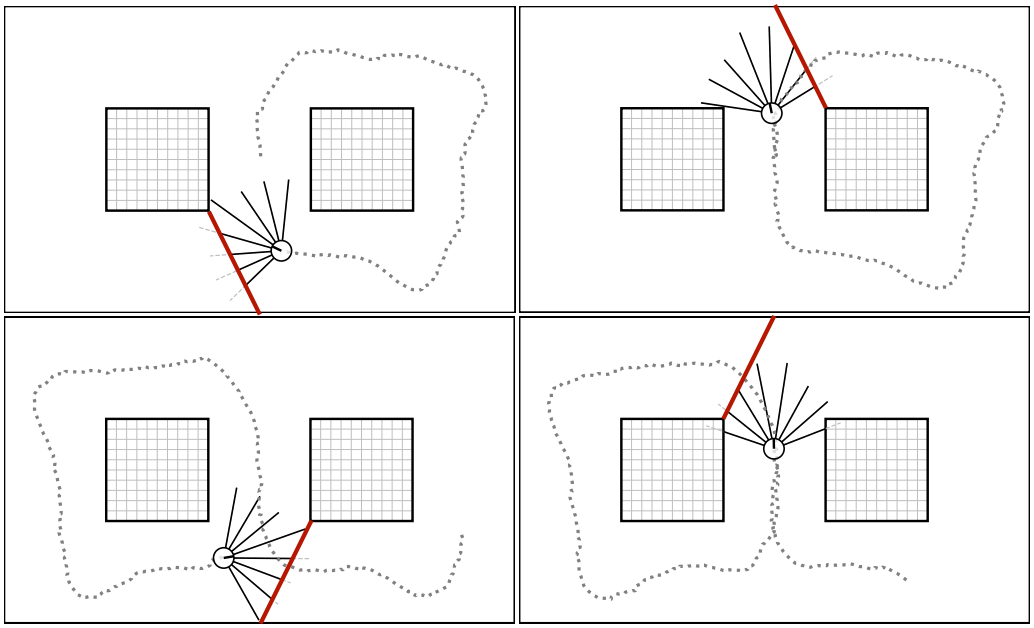


Figure 4.14: Generation of the 8-shape pathway with the addition of walls at the intersection points

Training data The ESN is trained using supervised learning, containing samples from the desired 8-shaped trajectory. Since the Braitenberg algorithm only aims to avoid obstacles, the robot is forced into the desired trajectory by adding walls at the intersection points as shown in figure 4.14. After generating the right pathway, the added walls are removed and the true sensor values are gathered as input. Gaussian noise is added to the position values of the robot at every time step in order to make the training more robust. Approximately 50,000 time steps were generated (equivalent to 71 complete 8-loops) and separated into training and testing sets.

Hyper parameters tuning The ESN was built with the python library ReservoirPy [198] with the hyper-parameters presented in table 4.1, column "Without context". The order of magnitude of the hyper-parameters was first found using the Hyperopt python library [20], then these were fine tuned manually. The ESN receives as input the values of the 8 sensors and output the next orientation.

Model evaluation The performance of the ESN has been calculated with the Normalized Root Mean Squared Error metrics ($NRMSE$) and the R square (R^2) metrics, defined as follows :

Parameter	Without context	With context
Input size	8	10
Output size	1	1
Number of units	1400	1400
Input connectivity	0.2	0.2
Reservoir connectivity	0.19	0.19
Reservoir noise	1e-2	1e-2
Input scaling	1	1(sensors), 10.4695 (cues)
Spectral Radius	1.4	1.505
Leak Rate	0.0181	0.06455
Regularization	4.1e-8	1e-3

Table 4.1: Parameter configuration for the ESN

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}}{\sigma} \tag{4.13}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{4.14}$$

where y_i, $\hat{y}_i$ and $\bar{y}$ are respectively the desired output, the predicted output and the mean of the desired output.

Reservoir state analysis In this section the reservoir states are analysed such as to inspect to which extent they form an internal and hidden representation of the memory. To do so, we use Principal Component Analysis (PCA), a dimensionality reduction method enabling the identification of patterns and important features of the processed data. PCA is carried out on the reservoir states for each position of the robot during the 8-shape trajectory. We continued the analysis by doing a classification of the reservoir states. We made the assumption that it is possible to know the future direction of the robot observing the internal states of the reservoir. This implies that the reservoir states can be classified in two classes: one related to the prediction of going left and the other related to the prediction of going right. Two standard classifiers, the KNN (K-Nearest Neighbors) and the SVM (Support Vector Machine) were used. They take independently as input, the reservoir state at each position of the bot while executing the 8-shape and predict the decision the robot will take at the next intersection (see figure 4.13). Since the classifiers are trained using supervised learning, the training data were generated in

the central corridor of the maze (yellow area in figure 4.12-Right), assuming that it is where the reservoir is in the state configuration in which it already knows which direction it will take at the next intersection. 900 data points were generated and separated into training and testing sets.

Experiment 2 : 8 Maze Task with Contextual Inputs

In this experiment, we fed the reservoir with two additional inputs that represent the next decision, one being related to a right turn (R) and the other to a left turn (L) (see figure 4.13). They are binary values, switched to a value of 1 only when the bot is known to take the corresponding direction. We thus built a second ESN with the hyper-parameters presented in TABLE 4.1, column "With context". The network is similar to the previous one, except that the contextual inputs are added with a different input scaling than the one used for the sensors inputs. During the data generation, the two additional inputs are set to 0 everywhere in the maze, except in the central corridor.

4.3.2 Results

Motor sequence learning

We first show that a recurrent neural network like the ESN can learn a rule-based trajectory in the continuous space, without an explicit memory or feedback connections. The score of the ESN is shown in TABLE 4.2 and the results for the trajectory predicted by the ESN are presented in figure 4.15 and in the top panel in figure 4.19. At each period of about 350 steps, a behavior or decision switch takes place, which is evident from the crests and troughs in the y-axis coordinates. It can be seen that the ESN correctly predicts the repeated alternating choice in the central arm of the maze. In addition to switching between the left and right loops, the robot also moves through the environment without colliding into obstacles.

Performance of the ESN for 50 simulations			
$NRMSE$		R^2	
Mean	0.0171	Mean	0.9962
Variance	5.4466e-06	Variance	1.0192e-06

Table 4.2: $NRMSE$ and R^2 score of the ESN with 8 inputs

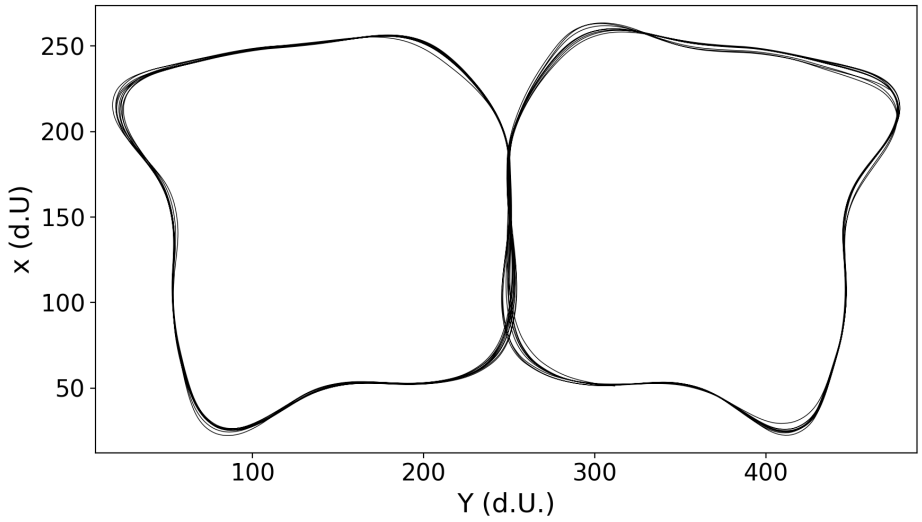


Figure 4.15: The trajectory of the robot following the 8-trace in the cartesian map.

Reservoir State Prediction

Next, we show that even a simple classifier such as SVM or KNN can observe the internal states of the reservoir and learn to predict the decision (whether to go left or right) of the network. The results of the predictions are presented in the top part of figure 4.16. As expected, there is a periodicity of choice in line with the position of the bot in the maze, showing that the classification is relevant. At each time step, both classifiers output the same prediction with a small discrepancy in time. The accuracy score obtained for both classifiers is 1. In the bottom part of figure 4.16, we can observe that the robot knows quite early which decision it will take at the next loop while we could expect that it would take its decision in the yellow corridor in figure 4.12. Here, we see that if the robot just turned right, the reservoir switches its internal state to go left next time only a few dozen time steps after. We tried the same classifiers but instead of the reservoir states as input, we used the sensors values. The results are shown in the figure 4.16. As expected, the classifications fail with an accuracy score of 0.57 for SVM and 0.43 for KNN; this randomness can be seen in both figures. Thus, we showed that by simply observing the internal states of the reservoir, it is possible to predict its next prediction. In essence, this is a proof of concept to show that second-order or observer networks, mimicking the role of the regions of the prefrontal cortex implementing contextual rules, can consolidate information linking sensory information to motor actions, to develop relevant contextual representations.

Since the state space of the dynamic reservoir is high-dimensional, using the Principal Component Analysis (PCA) on the states, we investigated if it is possible to observe sub-space attractors. The result for the PCA analysis is presented in figure 4.18, where PCA was applied for 5000 time steps, which corresponds to 7 8-loops. The figure shows two symmetric sub-attractors, which are linearly separable, that correspond to the two parts of the 8-shape trajectory.

Explicit rules with contextual inputs

Finally, we demonstrate that although the ESN can learn a motor sequence without contextual inputs, it is limited by its internal representation to learn more complex sequences which may require a longer memory. Adding contextual or explicit information about the rule (which we propose are representations developed by the prefrontal cortex over time) can then bias the ESN to follow any arbitrary trajectory as in 4.19. With the additional contextual inputs, the ESN is able to reproduce the standard 8 sequence (the performance is shown in table 4.3) but can also achieve more complex tasks by sending to it the proper contextual inputs. One ex-

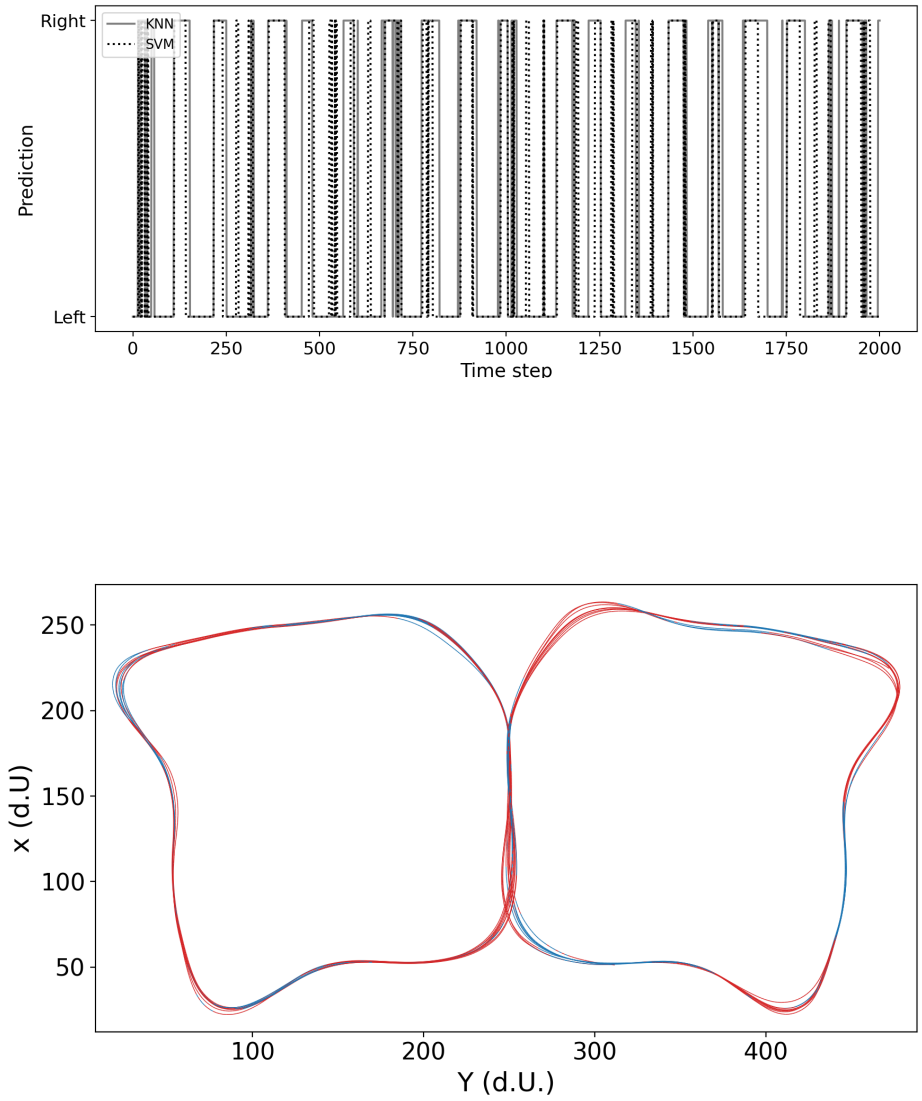


Figure 4.16: Prediction from sensors during 2000 time steps. Top figure shows the prediction of the KNN and SVM classifier, bottom figure shows the SVM prediction along the trajectory.

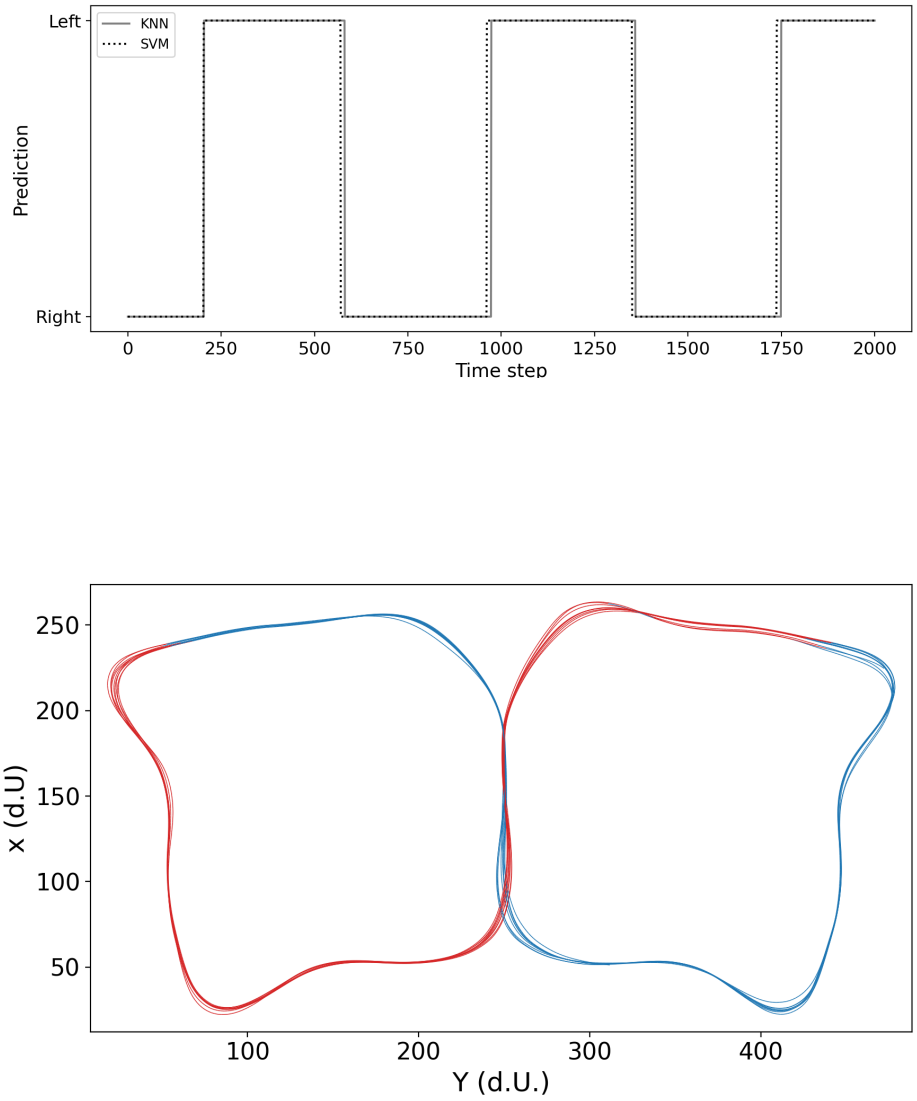


Figure 4.17: Prediction from reservoir state during 2000 time steps. Top figure shows the predictions of the KNN and SVM classifier. Bottom figure shows the SVM prediction along the trajectory.

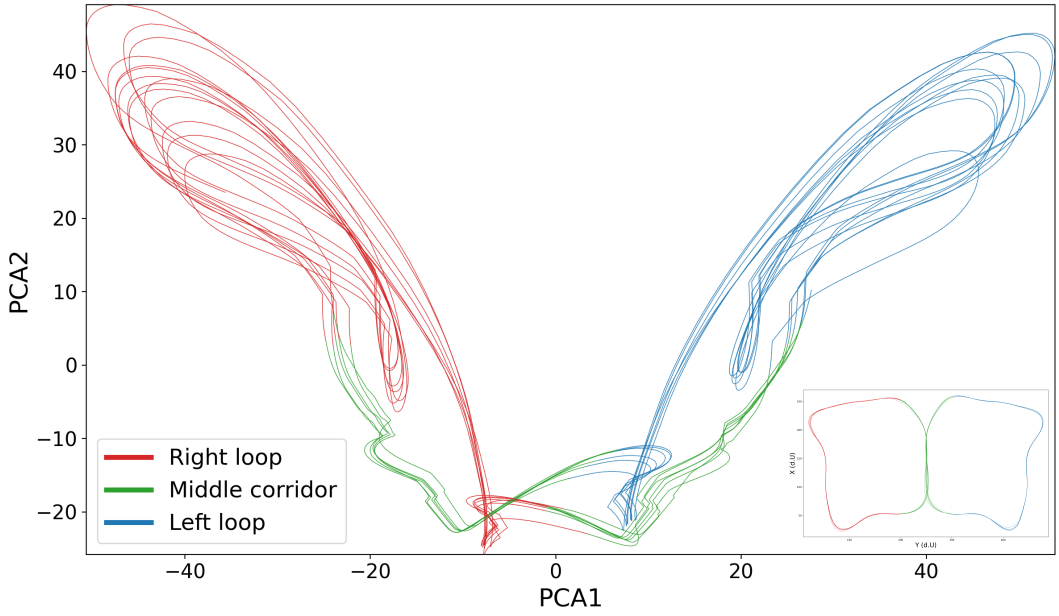


Figure 4.18: The first two principal components of the reservoir state space after applying PCA on the reservoir states. On the bottom right is the corresponding map of the positions of the robot in the maze.

Performance of the ESN for 50 runs			
$NRMSE$		R^2	
Mean	0.0050	Mean	0.9997
Variance	1.1994e-07	Variance	2.0220e-09

Table 4.3: $NRMSE$ and R^2 score of the ESN with the two additional contextual inputs

ample can be seen in figure 4.19: the top graph shows the positions of the bot while making the standard 8 sequence [ABABABAB...], the bottom one shows that the bot was able to achieve a more complex sequence [AABBAABBAABB...].

4.3.3 Discussion

Using a simple reservoir model that learns to follow a specific path, we have shown how the resulting behavior could be interpreted as an alternating behavior by an external observer. However, we've also shown that from the point of view of the model and in the absence of associated cues, this behavior cannot be interpreted as such. Instead, the behavior results from the internal dynamics of the reservoir (and the learning procedure we implemented). Without external cues, the model

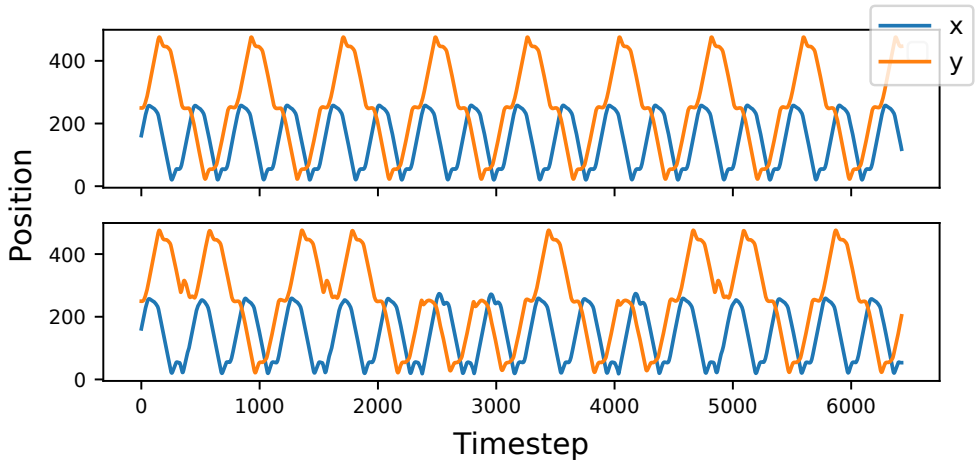


Figure 4.19: The coordinates of the agent for 7000 timesteps in the prediction phase. The plots in blue show the x axis coordinates while the ones in red show the y axis coordinates. The figure on top shows the results for the standard 8 sequence [ABABAB..], the figure at the bottom shows the results for a randomly generated sequence [AABAABBBABBAABAB], where 'A' is the left loop and 'B' is the right loop.

is unable to escape its own behavior and is trapped inside an attractor. Only the cues can provide the model with the necessary and explicit information that in turn allows to bias its behavior in favor of option A or option B.

From a neuroscience perspective, as developed in more details in [111], it can be proposed that the reservoir model in the first experiment implements the premotor cortex learning sensorimotor associations in the anterior cortex. In the first experiment, this is made by supervised learning in a process of learning by imitation. In a different protocol, this is also classically done by reinforcement learning, involving another region of the anterior cortex, the anterior cingulate cortex, manipulating prediction of the outcome. Whereas both regions of the anterior cortex are present in mammals, [111] reports that another region, the lateral prefrontal cortex, is unique in primates and has been developed to implement the learning of contextual rules and to possibly act in a hierarchical way in the control of the other regions. We have proposed an elementary implementation of the lateral prefrontal cortex in the second experiment, adding explicit contextual inputs as a basis to form contextual rules. It was accordingly very important to observe that it was then possible to explicitly manipulate the rules and form flexible behav-

ior, whereas in the previous case, rules were implicitly present in the memory but not manipulable.

This simple model shows that the interpretation of the behavior by an observer and the actual behavior might greatly differ even when we can make accurate prediction about the behavior. Such prediction can be incidentally true without actually revealing the true nature of the underlying mechanisms. Based on the reservoir computing framework which can be invoked for both premotor and prefrontal regions, we have implemented models which are structurally similar (as it is the case for that regions) and we have shown that a simple difference related to their inputs can orient then toward implicit or explicit learning as respectively observed in the premotor and lateral prefrontal regions. It will be important in future work to see how these regions are associated to combine both modes of learning and switch from on to the other depending on the complexity of the task.

4.4 Discussion

The approach towards the modeling efforts presented in this chapter was to investigate the basic constituents of cognitive control, using simple computational frameworks. In this respect, the contribution of this work are three fold :
In the first section, we illustrate the necessity of having multiple memory systems (working memory and episodic memory). As discussed in chapter 2, we highlighted the difference between the rodent and primate cortex as that of the absence or presence of the granular cortex. The limbic and agranular regions can be considered homologous for interpreting our model. In that aspect, we demonstrate that learning simple associations involves the striatum, while learning contextual associations may involve a basic version of a working memory system, and switching between contextual associations may involve the episodic memory system. We proposed earlier that some parts of the agranular cortex in rodents may play the role of regulating control. Following from this, what we show in our results is that a basic form uncertainty monitoring algorithm may be sufficient to describe change detection in the environment, and signal the need to explore (foraging).

In the second section, we show how adding a bias akin to a 'default behavior' to the same framework can better explain experimental behavioral results that use a more ecological setting than the restrictive T or Y maze. The question as to how the default behavior is overcome in absence of an explicit context in rodents - whether by creating a latent state representation of the context, or by simply coping with pre-learned associations remains unresolved for now and may need specific experiments to probe further. Nevertheless, our results strengthen our initial view of

cognitive control as the ability to over come default behaviors, an ability not special to primates.

In the third section, we use the reservoir computing framework which can be thought of as analogue for the motor and premotor regions, to show that a simple difference related to their inputs can orient then toward implicit or explicit learning. This is also consistent with the functional organization we have presented previously, with the motor and premotor connected to the preSMA in both rodents and primates. This experiment also highlights the question raised in the previous section, about the importance of explicit contexts for efficient manipulation of behavior.

The aim of this Chapter was two fold : 1. To confirm and illustrate our view of cognitive control as a gradient, through the computational modelling of increasingly complex tasks, 2. To investigate the role of Cognitive Control is resolving the interference that may arise from these different learning and memory systems. To this end, our work highlighted the importance of a "context" in switching between behaviors, and in the next chapter, we posit that the lateral PFC is particularly well suited for the elaboration of contextual behaviors.

In the next chapter, we show how the architectural addition of the lateral PFC, along with its hierarchical organization, allows for an elaborate representation and maintenance of contextual rules over time, thus endowing primates and humans with the ability to learn, and switch between abstract rules.

5 Cognitive control over contextual and abstract Rules

In Chapter 4, we explored the basic constituents of cognitive control, particularly concluding on the role of contexts in executing efficient cognitive control. While previously, we focused on studies and experimental results reported in the rodent literature, here we study the experimental paradigms used to probe more complex Cognitive Control in humans. As discussed in Chapter 2, the distinguishing feature of the primate brain is the granular cortex, in particular the lateral part of the prefrontal cortex. While the models presented in Chapter 4 focused on the role of the medial PFC (and more generally the limbic regions of the frontal cortex) in learning the values of goals and actions, this Chapter elaborates on the role of the lateral part of the prefrontal cortex (or the associative regions) in learning, representing and processing abstract rules. The interactions between the medial and lateral regions are discussed, with the latter being activated by the former for contextual control. Finally, while the previous chapter was about the mediation between different learning strategies an animal might use, this chapter aims to describe the PFC's role in the selection, maintenance and manipulation of learned behaviors.

Flat learning methods learn only one set of stimulus action mappings (referred to as a Rule Set, and several Rule Sets make up a Task Set). As a result, each stimulus corresponds to a single ideal action. This is limited as a model of human cognition since the best optimal action given a certain input typically relies on the current goal or context. While changing contexts, a flat learning agent would have to repeatedly override or relearn this information (called catastrophic forgetting in case of reversal learning). Biological agents though, must have the ability to cope with, and benefit from continuous, lifelong, online learning. A hierarchical architecture poses a natural solution to the storage of numerous task sets and adaptive switching between them in response to significant changes in the stimulus action reward contingencies. In such architectures, several rule sets are learned at the low-

est level (concrete rules) and the agent chooses when to transition between various rule sets at the upper level (abstract rules). At even higher levels of the hierarchy, i.e, more complex or abstract rules, the decision can be to switch between task sets. Thus, the role of rule and task sets is to provide protection from forgetting the established, previous learning.

To reiterate the distinction between concrete and abstract rules (presented in 1.5.1), in the simplest case, this means learning rules defined as associations between an object's properties and a direct response. Such rules can be called concrete, while more complex or abstract rules may involve the learning of second order relations on top of the first-order rules. The prefrontal cortex (PFC) is believed to provide the ability to contextualize concrete rules that leads to the acquisition of abstract rules [63]. Considering the number of contexts we encounter every day and the ease with which we select appropriate strategies for each, some relevant questions arise : How do we represent these strategies or rules and how do we determine which one is appropriate ? An important way of understanding how the PFC supports contextual learning and implements cognitive control is thus to understand how its representations are organized and manipulated [211].

There is sufficient evidence to suggest that the PFC is organized hierarchically [11] with more caudal areas learning first-order associations and more rostral areas putting them in context to facilitate learning of abstract rules. The hierarchy itself can be at two levels : a hierarchy between concrete and abstract rules, and a hierarchy of abstract rules, represented in increasingly rostral areas. The selection of the appropriate rule, depending on the context, is done by top-down modulation in the PFC, which underlies the ability to focus attention on task-relevant stimuli and ignore irrelevant distractors, in two ways in models : either as a result of weight changes in modulated pathways and predictions, or through activation-based biasing provided by a working memory system. Theoretically, despite agreement on the presence of an anterior-posterior (or rostral-caudal) gradient in the PFC, there is substantial ambiguity regarding the representational and/or processing demands that underlie this organization [169]. Two major hypotheses have been proposed :

The *information cascade hypothesis* [113]: The anterior-posterior gradient of lateral PFC is organized in a such a manner as to *when* a cue stimuli (or context) reduces uncertainty (i.e. provides information useful for) in the action selection process. In order to respond selectively to task signals that are temporally removed from the action selection process, anterior portions of the lateral PFC must be maintained for a long period of time (i.e. across multiple trials). On the contrary, it is hypothesized that posterior areas are receptive to both - cues that are important across trials and cues that arise in close temporal proximity to action selection (i.e. in the same trial).

The *levels of abstraction hypothesis* [13]: The anterior-posterior gradient of PFC is organized according to the level of abstraction (or hierarchical nesting) of the cues required to guide action selection. Thus, the processing of more abstract information about actions (such as collections of stimulus-response mappings) is selectively to anterior parts of the lateral PFC, whereas the processing of more specific information about actions is thought to be linked to posterior areas of the PFC (e.g. individual stimulus-response mappings).

These two hypotheses make similar predictions under many circumstances, and have been explored in a computational model - the Hierarchical Error Representation (HER) [3] model, which explains cognitive control in terms of the interaction between the dlPFC (dorsolateral prefrontal cortex) and the mPFC (medial part of the PFC). The dlPFC learns to maintain representations of stimuli that reliably co-occur with outcome prediction error and these error representations are used by the mPFC to refine predictions about the likely outcomes of actions. The error is broadcasted through the PFC in a bottom-up manner, and modulated predictions from top-down facilitate selection of an appropriate response. Thanks to its recursive architecture, this model, presented in more details below, can elaborate hierarchical rules on the basis of learning by weight updating, both to select pertinent stimuli and to map a representation inspired with principles of predictive coding [2].

In addition to its elegant recursive mechanism, proposing an original computational mechanism to account for the hierarchical structure of the PFC, the HER model is also very interesting because it proposes to decompose the functioning of the PFC between, on the one hand, the prediction of the outcome and the monitoring of the error of prediction and, on the other hand, the elaboration of contextual (and possibly hierarchical) rules to compensate errors. The context under this framework, is built from error representations in the dlPFC. This distribution of functions has also been reported between respectively the medial and lateral parts of the PFC [63], yielding more importance to the biological plausibility of the HER model. For these reasons, the HER model could be presented as a more elaborated and accurate model of the PFC, except for two points of discussion that we put forward here.

In the work presented here, we seek to answer specific questions about the nature of top-down modulation and selective attention, through the lens of hierarchical learning and representations. In the first section, we start from the implementation of the HER model with a reinforcement learning training signal for the gating mechanism instead of the back-propagated error signal reported in the original paper. We report the benchmark performance of this model on two different kinds of hierarchical tasks (corresponding to the information cascade and level of

abstraction hypotheses). Further, we use the model to study a task in which individual first-order rules can be learned alone or associated within specific contexts to form second-order rules. In the second section, we evaluate the performances of the HER model in the different cases in such a task, in which the model fails to perform, and compare them with a case where an attentional mechanism should be deployed to facilitate and orient its learning. As discussed in the concluding part, we observe that the attentional mechanism should be considered not only for the processing of information but also for the learning of rules, particularly in the hierarchical and contextual case.

5.1 The Hierarchical Error Representation (HER) Model

5.1.1 Methods

This section first summarizes the HER model algorithm and equations, as described in the original paper [3] and the subsequent section presents the tasks that we have chosen for our study.

Model Details : HER

Working Memory Gating

At each level of the hierarchy, external stimuli presented to the model may be stored in WM based on the learned value of storing that stimulus versus maintaining currently active WM representations.
External stimuli are represented as a vector **s**, while internal representations of stimuli are denoted by **r**. The value of storing the stimulus represented by **s** in WM versus maintaining current WM representation **r** is determined as :

$$\mathbf{v} = \mathbf{X}^{\mathbf{T}}\mathbf{s} \tag{5.1}$$

where **X** is a matrix of weights associating the external stimuli (**s**) with corresponding WM representations (**r**).
The value of storing stimulus $s_i(v_i)$ is compared to the value of maintaining the current contents r_j of WM (v_j) using a softmax function :

$$\text{probability of storing} s_i = \frac{(exp^{\beta v_i} + bias)}{(exp^{\beta v_i} + bias) + exp^{\beta v_j}} \tag{5.2}$$

Outcome Prediction

Following the update of WM, predictions regarding possible responses and outcomes are computed at each hierarchical layer, using a simple feedforward network :

$$\mathbf{p} = \mathbf{W}^{\mathbf{T}}\mathbf{r} \tag{5.3}$$

where **p** is a vector of predictions of outcomes and **W** is a weight matrix associating **r** and **p**

Top-Down Modulation

Beginning at the top of the hierarchy, predictions are used to modulate weights at inferior layers and modulated predictions are computed, as shown with the red arrows in figure 5.1.
For a given layer, the prediction signal **p'** additively modulates stimulus-specific predictions **p** generated by the lower layer. In order to modulate predictive activity, **p'** is reshaped into a matrix **P'** and added to **W** in order to generate a modulated prediction of outcomes :

$$\mathbf{m} = (\mathbf{W} + \mathbf{P}')^{\mathbf{T}}\mathbf{r} \tag{5.4}$$

These modulated predictions are then used to modulate predictions of additional inferior layers (if any exist)

$$\mathbf{m} = (\mathbf{W} + \mathbf{M}')^{\mathbf{T}}\mathbf{r} \tag{5.5}$$

Response Selection

Actions are learned as response-outcome conjunctions at the lowest layer of the hierarchy. In fact, a weakness of functional description of the model is that these concrete rules are simply represented at the lowest level of the hierarchy, while according to our functional description, these rules must be learned or represented in the premotor cortex, and thus not be part of the prefrontal hierarchy per se, a distinction we clarify in our proposed model. To select a response, the model

compares the modulated prediction of correct feedback to the prediction of error feedback, for each candidate response :

$$u_{response} = m_{response/correct} - m_{response/error} \tag{5.6}$$

This is then used in a softmax function to determine a response :

$$Prob(u_i) = \frac{\exp^{\gamma u_i}}{\sum \exp^{\gamma \mathbf{u}}} \tag{5.7}$$

Bottom-up Process

Following the model's response, it receives feedback regarding its performance and two error signals are computed at the bottom most hierarchical layer, one comparing the unmodulated predictions to the outcome :

$$\mathbf{e} = \mathbf{a}(\mathbf{o} - \mathbf{p}) \tag{5.8}$$

and another comparing the modulated predictions to the outcome :

$$\mathbf{e} = \mathbf{a}(\mathbf{o} - \mathbf{m}) \tag{5.9}$$

where **o** is the vector of observed outcomes and **a** is a filter that is 0 for outcomes corresponding to unselected actions and 1 everywhere else.

The outer product of the first error signal and the current contents of the WM at the bottom level is used as the feedback signal for the immediately superior layer where this process is repeated (figure 5.1).

$$\mathbf{O}' = \mathbf{r}\mathbf{e}^{\mathbf{T}} \tag{5.10}$$

Effectively, at the second layer, the outcome matrix is a conjunction of stimuli, actions and outcomes. This matrix is reshaped into a vector **o'** and used to compute the prediction error at the superior layers :

$$\mathbf{e}' = \mathbf{a}'(\mathbf{o}' - \mathbf{p}') \tag{5.11}$$

Weights Updating

The second error signal is used to update weights within the bottom-most hierarchical layer, it updates the weights connecting the WM representation to prediction units (**W**), as well as weights in the WM gating mechanism (**X**):

$$\mathbf{X_{t+1}} = \mathbf{X_t} + (\mathbf{e_t^T W_t} \cdot \mathbf{r_t})\mathbf{d_t^T} \tag{5.12}$$

An eligibility vector **d** is used instead of the stimulus vector **s**. When a stimulus i is presented, the value of d_i is set to 1, indicating a currently observed stimulus and at each iteration of the model, **d** is multiplied by a constant decay parameter indicating gradually decaying eligibility traces.

The above equation uses a backpropagated error to train the associative weights **X**, a learning mechanism considered neurally implausible.

New learning rule:
We have proposed (and use for the rest of this work), a more biologically plausible training of the gating mechanism, using a scalar reinforcement learning signal. On each simulated trial, the model receives feedback regarding its performance : error or correct. In a reinforcement learning framework, such feedback constitutes a binary reward signal which takes the value of 1 for correct feedback, and 0 for error feedback. In RL, learning is driven by the difference between a received reward and a predicted reward; however the HER model learns predictions of multiple possible outcomes based only on the likelihood of observing the outcome and without regard for its affective valence. In order to model reward predictions, the output of a softmax function for the selected behavior is used as a proxy for reward prediction, with a temperature parameter set to 5 in the simulations.

$$\text{Value} = \frac{\exp^{\gamma u_i}}{\sum \exp^{\gamma \mathbf{u}}} \tag{5.13}$$

When correct feedback is highly probable, the output of the softmax function will be close to 1 for the response associated with receiving correct feedback. Following feedback, a scalar error term is calculated as the outcome feedback (correct/error) minus the proxy reward prediction.

$$\delta_t = reward_t - Value_t \tag{5.14}$$

Weights in the WM mechanism are then updated according to :

$$\mathbf{X_{t+1}} = \mathbf{X_t} + \alpha_\mathbf{t} \mathbf{r_t} \mathbf{d_t^T} \tag{5.15}$$

Finally, the prediction weights in the model are updated according to

$$\mathbf{W_{t+1}} = \mathbf{W_t} + \alpha(\mathbf{e_t r_t^T}) \tag{5.16}$$

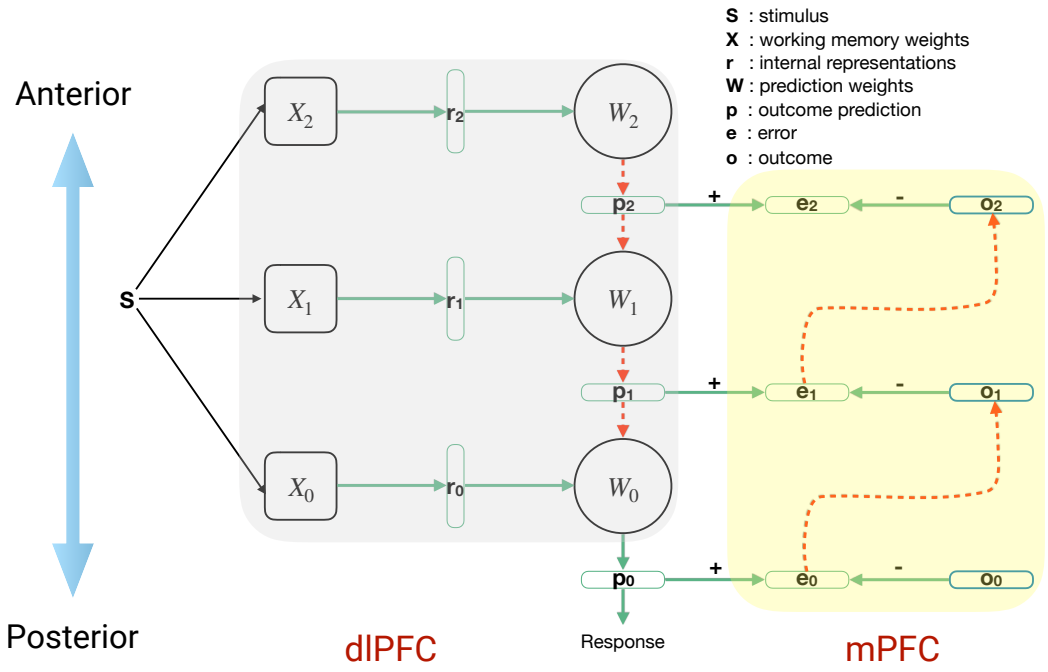


Figure 5.1: Model Schematics : Figure adapted from [3]

5.1.2 Tasks

12AX Continuous Performance Task

Stimuli are presented beginning with a context cue (1 or 2), followed by a pattern cue (A or B) and then with a target cue (X or Y). The context cue indicates which pattern cue - target sequence is valid (AX or BY). Sequences of stimuli are organized as inner and outer loops. This task is an extension of the AXCPT task. It is an example of a learning paradigm, where the hierarchy is present in the form

of temporal abstraction (the higher level context cue is to be maintained until a valid target sequence is observed). It tests the working memory of the subjects.

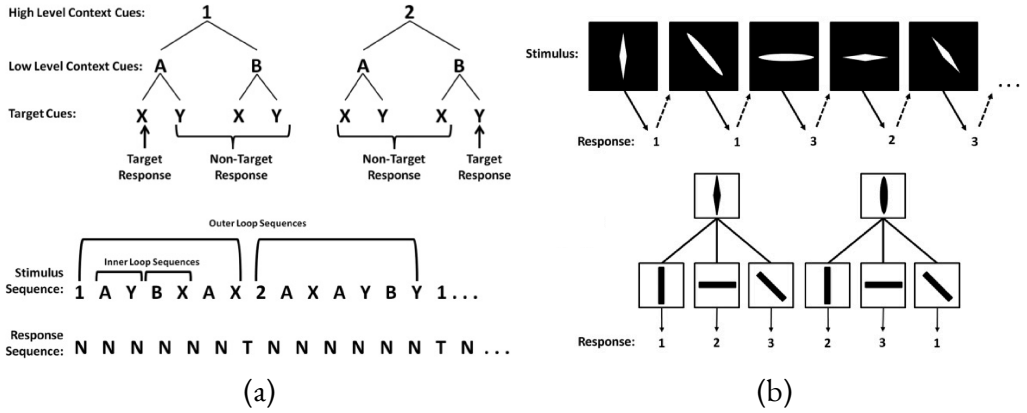


Figure 5.2: Task Schematics : **(a)** 12AX continuous performance task **(b)** Hierarchical categorization task, Figures taken from [3]

Structured Task

In the notional task, as depicted in figure 5.2, each compound stimulus consists of two dimensions, shape and orientation, with dimensionalities of two and three respectively. This results in a task with three possible responses and two response mappings. It should be noted that structured tasks of this sort can be solved using a generic backpropagation network with a single hidden layer. Although such networks are not hierarchically structured, they are able to learn hierarchical tasks through learning conjunction of features, represented by activity in hidden units, which can then be used to generate appropriate responses.

Hierarchical Categorization Task 1

In the hierarchical task reported by Badre, an individual stimulus conjunction consisted of one of 3 shapes, at one of 3 orientations, inside a box that was one of 2 colors, for a total of 18 unique stimuli (3 shapes x 3 orientations x 2 colors). The arrangement of response mappings for these stimuli was such that a second order relationship existed. In the context of one colored box, only the shape dimension

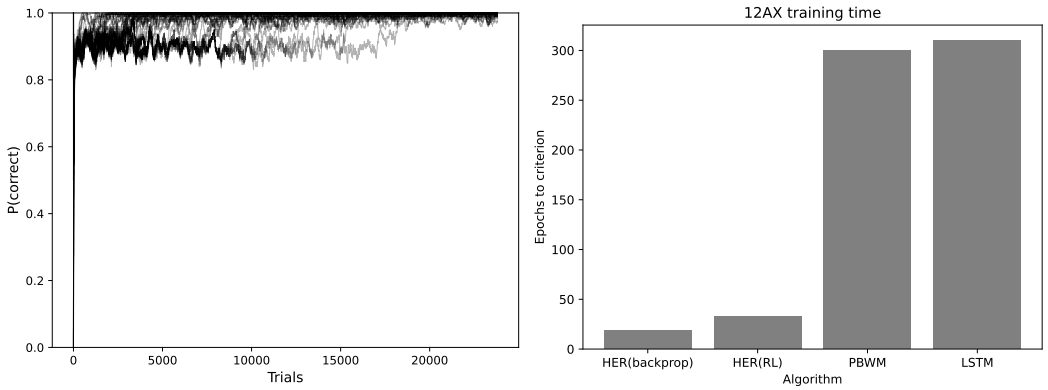
was relevant to the response, with each of the 3 unique shapes mapping to one of the 3 button responses regardless of orientation. Conversely, in the context of the other colored box, only the orientation dimension was relevant to the response. Thus, this task permitted learning of an abstract conditional rule that specified how one dimension (color) determined which of the other dimensions (shape or orientation) would provide a context for selecting a response.

5.1.3 Results : Benchmark Performance

Simulation 1 : 12AX Continuous Performance Task

The HER model was simulated on a version of the 1-2AX task as described in O'Reilly and Frank (2006) in which each outer loop consisted of one to four inner loops and the probability of observing a valid sequence for each inner loop was 0.25. There were six inputs to the model, corresponding to the six relevant cues in the task. At each cue, the model made a response to indicate whether the current sitmulus was a target (or not). Feedback to the model indicated correct or incorrect performance. One hundred simulations of the HER model performing the task were conducted for 4000 outer loop sequences (equivalent to 160 training epochs in O'Reilly & Frank, 2006, where 1 epoch was 25 outer loops, or approximately 24,000 individual cue presentations). The model was considered to have successfully learned the task on the first of 1000 consecutive cue presentations (approximately seven epochs) in which no response errors were made. The more lenient criterion as described in O'Reilly & Frank, 2006 was two consecutive epochs with no errors (or approximately 300 cue presentations).

Of the 100 simulations, the model successfully met criterion 83% of times. On average, the number of cue presentations to criterion was 8961.5 trials (approximately 61 epochs). According to the other criterion, on average, the number of cue presentations to criterion was 4867.57 trials (approximately 33 epochs)



(a) (b)

Figure 5.3: **(a)** Percent correct trials for 100 simulations of the HER model on the 12AX task. Values reflect the running average for a moving 200 trial window. **(b** Performance of the HER model in relation to simulations of the backprop version, PBWM and LSTM models using the criteria of 2 epochs without error. The RL version of the model, as reported here takes more time to converge than the version trained using backprop.

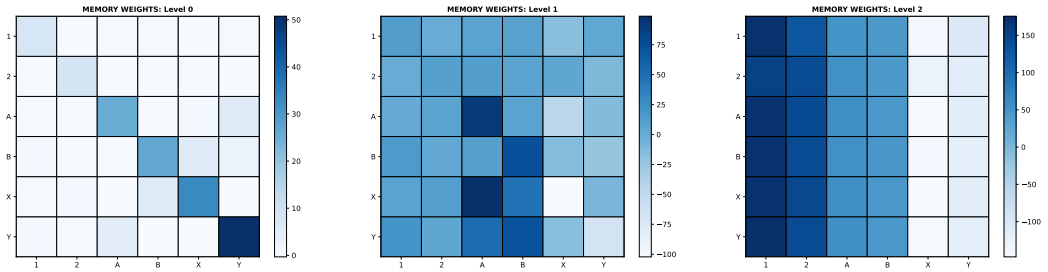
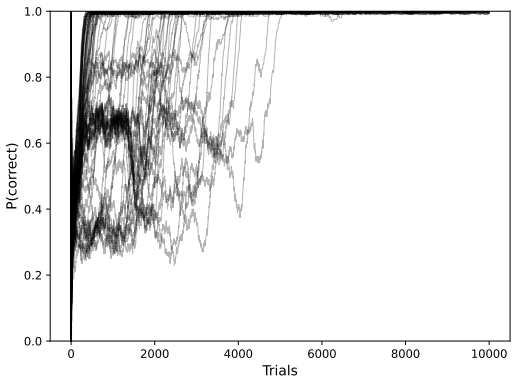


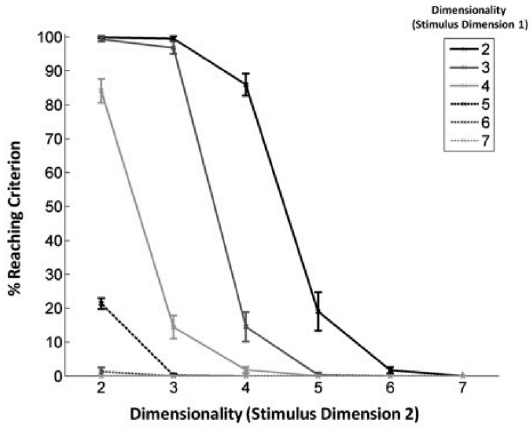
Figure 5.4: The memory weights X (averaged over 100 simulations) in the gating mechanism reflect the value of maintaining versus updating WM representations following learning. At the lowest hierarchical layer, weights are confined to the diagonal, indicating the mechanism tends to update WM on each trial, with weights corresponding to X and Y particularly strong indicating that these are the representations updated every time they are presented. In layer 2, the weights reflect the tendency to maintain representations of context variables A and B, rather than gating in X and Y representations, as seen by the positive off diagonal weights. Finally, weights in layer 3 strongly favor the representation of 1 and 2, regardless of what stimulus is presented.

Simulation 2 : Structured Task

The model was simulated performing the task for 4000 trials, and 100 simulations were performed. To assess performance, the model was considered to have completed learning the task on the first of 1000 consecutive trials in which no errors were committed. The results are show in figure 5.5 (a). The more complicated version of this task, in which one dimension provides the context for which one of the other dimensions must be attended to is unable to meet the convergence criteria.



(a)



(b)

Figure 5.5: **(a)** The model's performance on the structured task described. **(b)** As reported in the HER model, increasing the dimension of the stimulus leads to a significant inability of the model to converge. Hence, the more complicated version of the structured task (Hierarchical Categorization Task 1) does not meet the convergence criteria.

5.1.4 Discussion

In the predictive coding framework, top-down processes provide predictions from superior hierarchical levels to inferior levels, while residual prediction errors, ie, input that cannot be accounted for by the predictions supplied by top-down processes, are carried from inferior levels to superior levels. At a single level, the HER model suggests that error signals computed in the mPFC can be used to train representations of the error signal in dlPFC. Error representations learned by dlPFC are associated with task stimuli that reliably precede error prediction signals generated by mPFC such that, on subsequent stimulus presentations, error representations maintained in dlPFC may be deployed to reduce prediction errors in mPFC. Previous functional imaging results have suggested a correspondence between medial and lateral PFC hierarchies, arguing that medial PFC provides motivational signals, while lateral PFC provides selection, that is cognitive signals. The HER model differs from this proposal in that rather than motivational signals, the mPFC provides prediction error signals that train the lateral PFC regarding what information must be maintained for successful task performance. The representation scheme proposed by the HER model suggests that individual neurons in lPFC should code for components of a distributed error representation, with single units signaling the identity and likelihood of observing a particular error. The working memory activations constitute a de facto representation of inferred states and thus provide a context dependent pattern of activity that minimises prediction error.

Because of the specificity of representation at each hierarchical layer, the HER model is unable to generalize associations learned in one context to a novel one. Once representations are formed, prediction related activity in mPFC may be sufficient to elicit associated activity in dlPFC that may then be dynamically mapped to external stimuli. It is unlikely that detailed information carried by the error signal in the HER model is used to train a gating mechanism as implemented in real brains. Rather, it is more likely that such training involves a scalar signal, typically associated with dopaminergic activity.

5.2 Integrative Model

A hierarchical extension of the PBWM model [150] [77] proposes that hierarchical control can arise from multiple nested frontostriatal loops (loops between the PFC and the BG). The system adaptively learns to represent and maintain higher order information in rostral regions which conditionalize attentional selection in more caudal regions. All the adaptations of the HER model are made through learning

by weight modifications, whereas the property of working memory of the PFC, as it is for example exploited in the PBWM model, is often presented as a key mechanisms for its adaptive capabilities. These two mechanisms reflect two very different ways of learning: one purely cortical and slower way of building representations on the cortical surface on the longer term, and the other based on subcortical interactions that learn when to gate information (using the cortical representations) on a faster timescale. These two mechanisms on the surface seem incompatible, and difficult to compare. An important question is consequently to determine up to which point working memory and attentional modulations are necessary for the learning of hierarchical rules in cognitive control. Through a careful examination of these mechanisms, we conclude that they in fact are complementary, and we propose a model, and an experimental paradigm that is able to reconcile these two views of learning into a an integrative framework.

One way to learn efficiently in complicated surroundings is to use selective attention, which can narrow down the task's complexity. This means that attention is focused on a few environmental aspects that are relevant to the task at hand, while others are ignored or generalized over. By doing this, the number of stimulus configurations of different states that need to be considered is reduced. However, attention should only be directed to dimensions of the environment that are important for the task, i.e., dimensions that predict reward, in order to provide a suitable representation of the task's state for the learning process. It is not always clear which dimensions are relevant to a particular task and may need to be learned through experience, a mechanism thought to be implemented by the use of an output gate.

Output Gating

Working memory may not always contain information that is pertinent to behavior at any given time. Instead, it is also adaptive to regulate *which* WM representations can affect attention and action selection, and *when*. Such selection from inside the working memory, or singling out of the WM representations is resource demanding and PFC dependent. A possibility is that selection from within WM can be conceived of as a gating function, similar to that outline for WM updating. According to this perspective, an output gate can regulate the information flow within the WM, between an actively maintained but inert state, to one that is capable of exerting a top down influence on behavior. To put it another way, when the output gate is closed, information would be maintained, but not have any modulatory top-down effect on behavior. Conversely, when the output gate is

open, the maintained information would provide a top-down contextual signal (illustrated in figure 5.6). This issue of selection from within WM by output gating has been proposed as a solution by many computational models [].

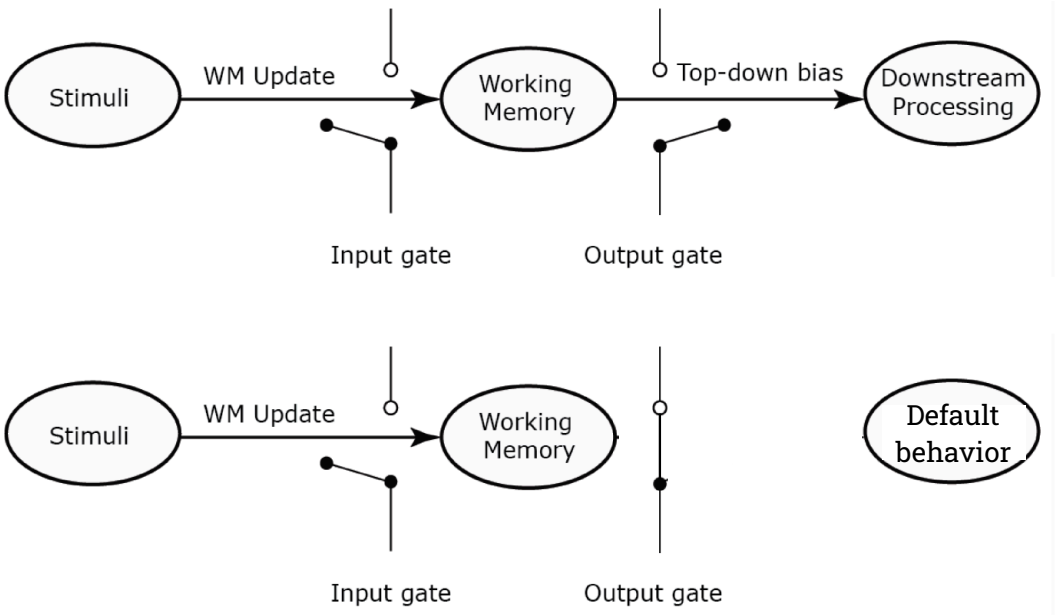


Figure 5.6: Illustration of output gating : access to WM is controlled via the operation of an input gate that determines whether a stimulus is updated into WM. An output gate controls whether or not information within WM can influence behavior. Figure adapted from [21]

5.2.1 Methods

Model

In the model with output gating and selective attention, instead of using modulated predictions P' from the superior layer, additively in the lower layer, the agent uses the prediction mapping, to rather select which stimulus dimension needs to be attended to, out of the ones maintained at the lower layer $r_{0,i}$

$$p_{m,i} = max(P'^{T} r_{0,i}) \tag{5.17}$$

These predictions are then used in a softmax function to determine which stimulus dimension (or rule set) will finally be selected to determine the response at the lower layer

$$Prob(r_{0,i}) = \frac{\exp^{p_{m,i}}}{\sum \exp^{\mathbf{p_{m,i}}}} \tag{5.18}$$

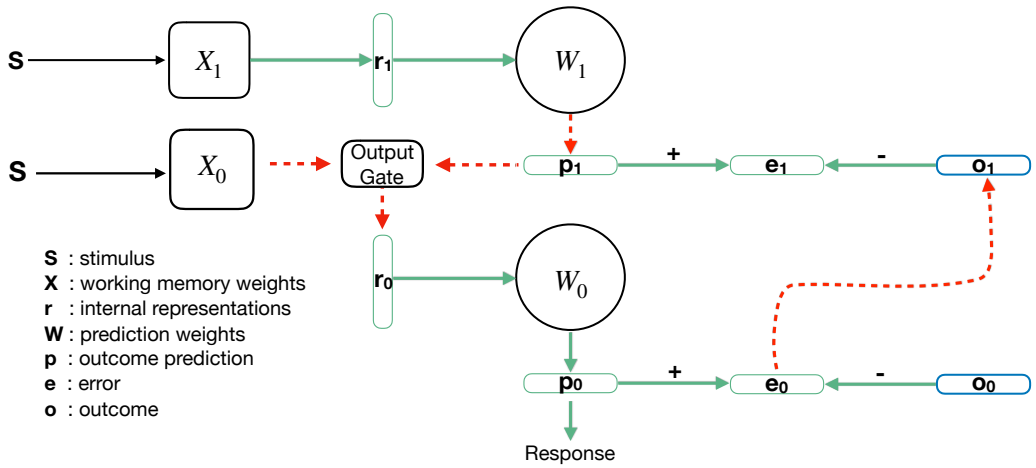


Figure 5.7: The modified model with output gating from layer 1. The gating weights in layer 1 (X_1) learn over time to gate the context into r_1. The selected prediction units from layer 1 (p_1) are then used to make a decision on which value of the stimulus s is gated into r_0 (the output gate).

5.2.2 Task

Hierarchical Categorization Task 2

To design our task, we consider the framework introduced by Koechlin [112] which is composed of three subtasks where the stimuli are letters having three dimensions: color (red, green or black), case (upper or lower) and sound (vowel or consonant). In the first subtask (Block 1 in figure 5.8(b)), black color indicates to ignore the stimulus and green color indicates to discriminate the case (rule T1: left button for upper, right button for lower). In the second one (Block 2 in figure 5.8(b)), black color indicates to ignore the stimulus and red color indicates to discriminate the sound (rule T2: left button for vowel, right button for consonant).

The third one (Block 3 in figure 5.8(b)) is a random mix of trials from the other two blocks. This framework is interesting because, whereas rules T1 and T2 in blocks 1 and 2 require the subject to attend to a single dimension of the stimulus, block 3 requires to pay attention to both and to decide which rule to apply based on the third (contextual) dimension. Let us also mention here that, while there is no apparent difficulty with such tasks, it is actually harder than it appears depending on the way a task is learnt. During block 1, one can either learn the rule : "green means case and black ignore" or the rule: "black ignore, else case". The same is true for block 2 with sound. If we now consider block 3 and depending on how a subject learnt the first two blocks, she may succeed or fail immediately. In this latter case, this means block 3 cannot exploit previous learning and has to be (re)learnt.

The original task was cued by instruction and corresponding performances were reported in the paper [112]. Here, we wish to explore the inherent capability of a model to learn an abstract and hierarchical rule task without instructional cues, as in the paradigm reported by [13] and also to consider how the hierarchy can be learnt, depending on how information is represented in the model. We used two types of learning paradigms for the simulations : the first paradigm in which rules T1 and T2 were learned one after the other, and the performance of the model was then tested on random trials interleaved from rule T1 and T2 (to say it differently, we apply successively block 1, 2 and 3). In the second paradigm, an entire abstract rule that we call T3, corresponding to the selection on rules T1 and T2 depending on the contextual cue 'color' was directly learned (block 3 applied first) and performance of the model was subsequently tested on rule T1 and T2 (blocks 1 and 2). In the next section, we report performances observed with the HER model and with an adapted version that we propose subsequently.

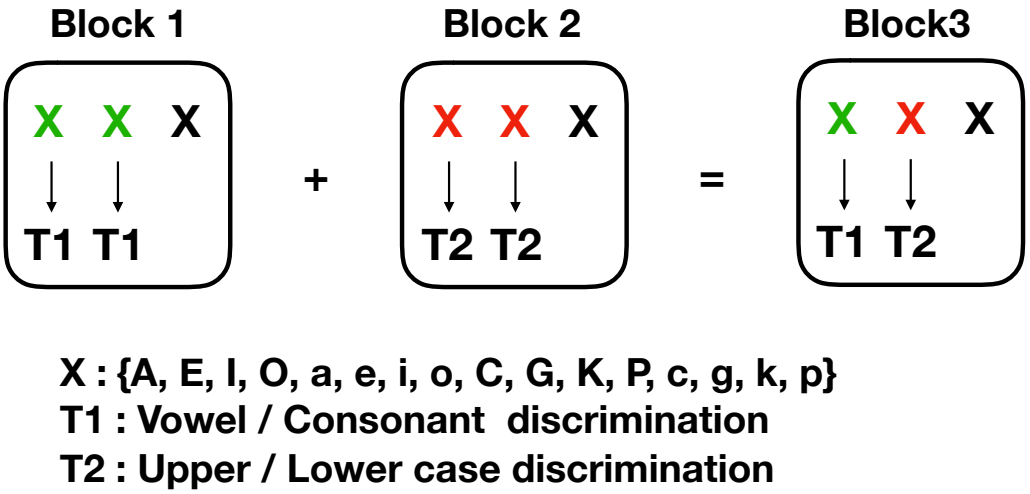


Figure 5.8: Task Schematics, as described in the section 5.2.2

5.2.3 Results

Simulation 1 : Shaping with original model

We have first studied how the HER model, as it has been designed (cf section 5.1.1), can address the tasks defined above, under the two mentioned paradigms (cf section 5.1.2). Due to the design of the HER model, each layer can only map or process one stimulus value, thus requiring as many layers as there are stimulus dimensions. The mapping in the model is also highly sensitive to the stimulus dimensions relative to one another, particularly higher-dimensional stimulus are preferentially mapped onto the lowest hierarchical layer. This rests on the assumption that stimulus dimensions better able to predict and reduce uncertainty about the response are mapped to lower layers.

This may not always be the case in real life situations though. We often have to adapt and generalize the same rules over several different contexts. In the task we consider as well, the context is determined by the color, which has 3 possible values - one of which always maps to the same response (to ignore) and the other 2 determine the response based on other stimulus dimensions.

Learning Curves

Performance observed for the first and second learning paradigms are reported in figures 5.9 (a) and (b) respectively. We see in the figure 5.9 (b) that due to its hier-

archical structure, when there is an underlying abstract rule to learn (rule T3), the model is able to use the hierarchical information to acquire the rule while retaining performance in each of the sub-rules (Rule T1 and T2). It does so by monitoring an "error of errors" at each hierarchical layer, broadcasting this error to superior layers (bottom-up processing) that put it in context with the stimulus feature being attended to and finally sends this prediction information to the lower layers (top-down modulation) which are able to then select the appropriate response. In the figure 5.9 (a), we show that when the composite rules are first learnt sequentially, the model is not able to compose them into a single rule, but instead has to relearn its representations to reach optimal performance.

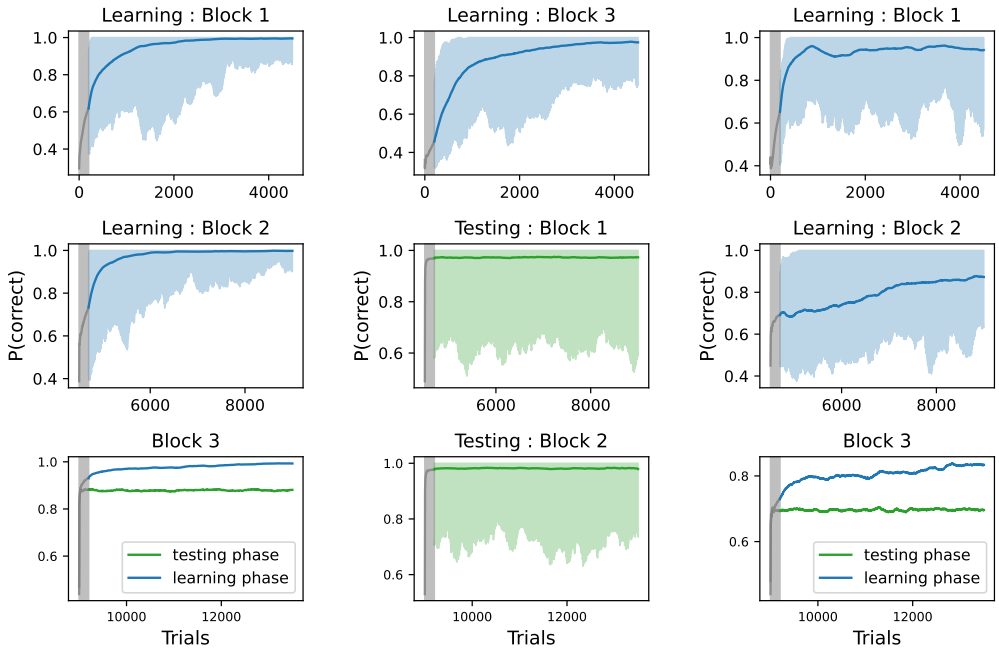


Figure 5.9: Performance of the model with 3 layers for the two paradigms **(a, b)**, plotted as an average over 100 runs, only for the runs that reached convergence criteria. The convergence criteria was defined as having a performance greater than 85% in the last 200 trials. **(c)** Performance for the model with 2 layers on the first learning paradigm.

Next we show that due to the design of the model, a task which has only one level of hierarchy, such as the one considered here, can not be learnt with a model with 2 layers. In figure 5.9 (c) we see that with 2 layers, the model is able to learn the subparts of the rule (rules T1 and T2), but performance on the composite rule T3 saturates at 80%. By exploiting the gating mechanism, each sub-rule can be learnt individually by gating the 2 relevant feature dimensions at the 2 layers (color, vowel/consonant for rule T1 and color, lower/upper case for rule T2). However, in the third rule T3 when the 2 relevant features change from trial to trial to determine the correct response, the model fails to learn, since the contextual stimulus features don't provide top-down information about "which" other stimulus feature to attend to at the lower layer.

Gating weights

In the model, the gating weights determine both, when to update or maintain a stimulus feature, and also which of the stimulus features is to be gated. We observed the adjusted weights after each rule that is learned. In the first block, vowel, consonant and black have high values of getting updated at the lowest layer, while in rule T3 all the "lower level" cues have high values of getting updated. In such a case, there is again competition between which one of them to gate, and both can win with close probabilities, in the absence of any information from the superior layers. Depending on what is gated into the top two layers, any of those mappings could emerge.

Prediction weights

The prediction weights at layer 0 are Stimulus-Action-Outcome conjugations and the gating mechanism determines which stimulus and in turn which action-outcome association is to be selected. The selected associations are then modulated by superior layers and used the determine the response. At layer 1, the prediction errors of layer 0 are contextualized to make SxSxAxO conjugations and so on.

In the task considered for all our simulations, there are 5 concrete rules or S-A-O predictions to learn : Black - Action3, Vowel, Lower case - Action1 and Consonant, Upper case - Action2 (figure 5.8 (b)). In figure 5.10, we present examples of how a model with 3 layers selects a response by additive prediction modulation. We observed that elaborating a mapping between the stimulus and what is gated into the internal representation ($\mathbf{r}$) at different layers could be done in different ways, including randomly, as long as these mappings led to orthogonal and mu-

tually exclusive activations of predictions (in **W**). For example, in figure 5.10 (e), in Block 2, the color red was not gated into the internal representation, but the random gating of the other 2 dimensions still led to an appropriate modulated prediction that could initiate the correct response.

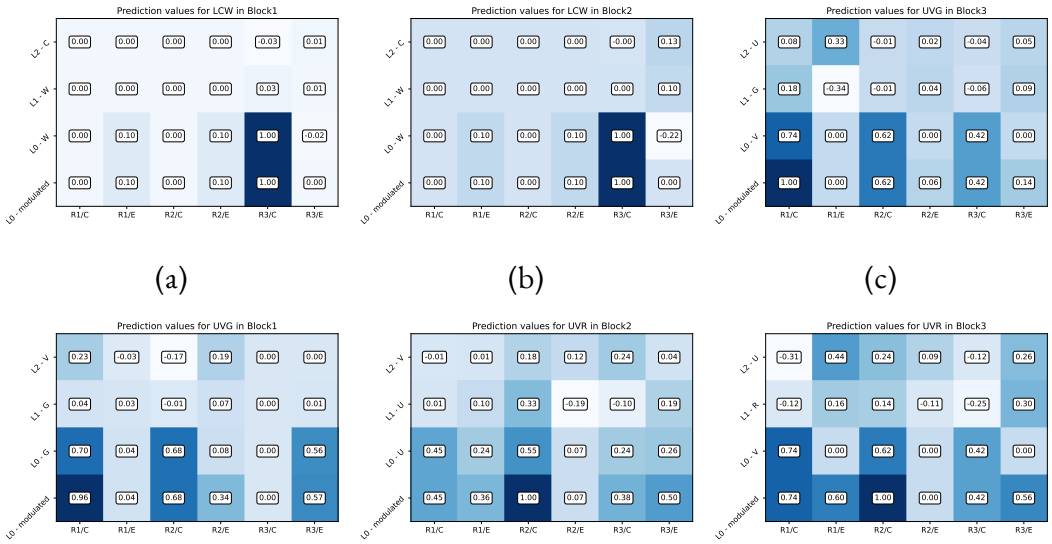


Figure 5.10: Examples of how the model solves different cases of stimuli. The matrix shows the prediction values at different layers (first 3 rows), given the internal representation of the stimulus, and how they are modulated additively (row 4) to give the final Action-Outcome predictions that are used for response selection. **a, b** show the case when the stimulus is black, in rules 1 and 2 respectively. **d, e** show the case when the stimulus is Green, Vowel (rule T1) and Red, Upper case (rule T2). **c, f** show the case for Green, Vowel and Red, Upper case in rule T3

Simulation 2 : Integrative model

To explain the deficit of attentional mechanism in the HER model, and illustrate the advantage of our proposal, we performed some simple simulations. The model was trained individually on the two discrimination tasks ie, on the two concrete rules (T1 - vowel/consonant and T2 - lower/upper case), to obtain prediction weights or Stimulus-Action-Outcome associations as in figure 5.11 (b). We tested

the ability of the HER model with 2 layers, to use this information and contextualize it to learn the abstract rule. The bottom layer of the model was initialized to the predictions previously learned and moreover, it was "frozen" such that no learning happened at this level, implying that these behaviors were rigid. At the upper layer, the gating weights were biased to update the internal representation with the context, which was the color in this case, implying saliency to previously unattended cues. As expected, the model failed to learn the abstract rule with these modifications. With the modified model, we used the same protocol ie the bottom layer was kept frozen, and there was a bias added to the upper layer to encourage gating of the color. However, instead of an independent gating at the bottom layer, we included an output gating from the upper layer, which used the prediction errors at the upper layer to select which stimulus dimension was going to be gated into the bottom layer (figure 5.11 (a)). To put it more generally, the bottom layer was responsible for response selection while the upper layer was responsible for action-set selection through targeted attention (cf [63] for more details about the structuring concept of action-set and its role in PFC information processing). Our modified model achieved optimal performance fairly quickly, as shown in 5.11 (c).

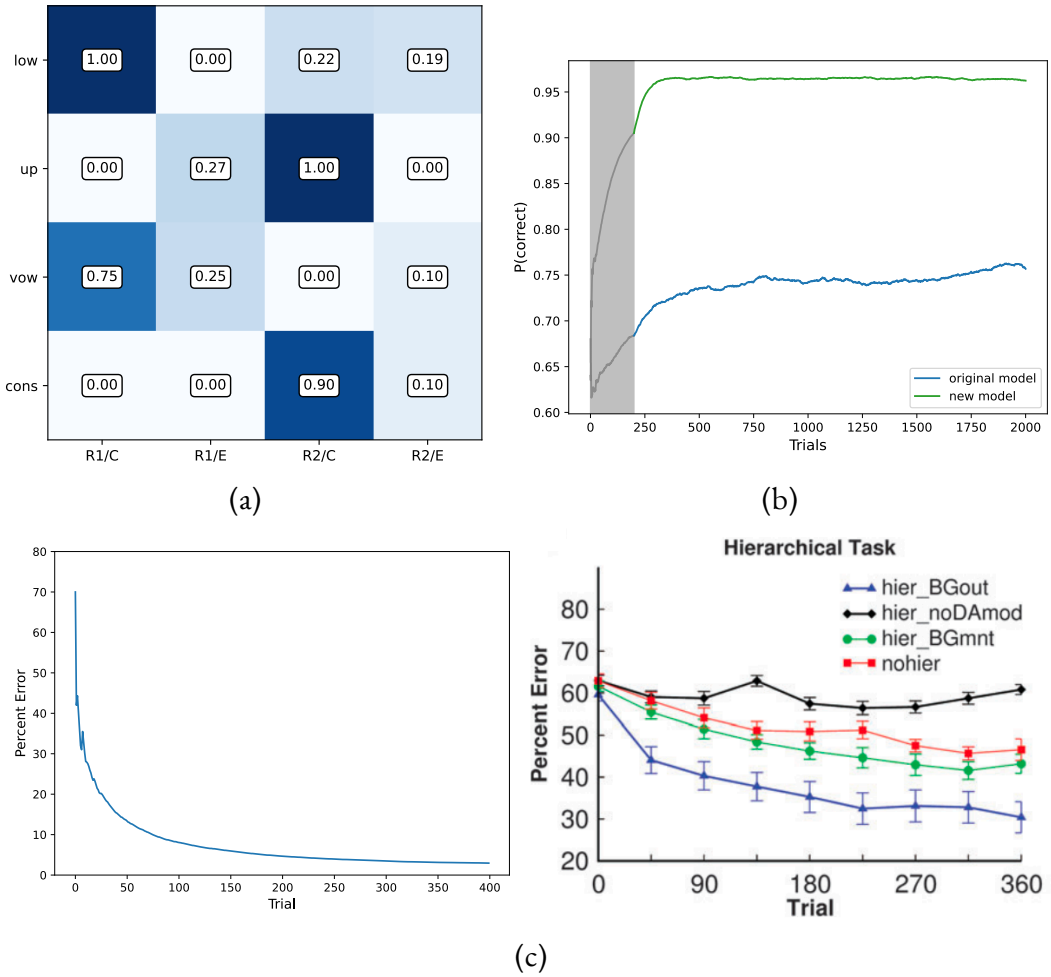


Figure 5.11: **(a)** Prediction weights (W_0) for the concrete rules at layer 0. These weights are pre-learned by training the model with rules T1 and T2, independently. **(b)** Performance of the original model compared to the modified model over a 100 runs, when layer 0 is fixed to the weights in figure (b) and only layer 1 prediction weights (W_1) and gating weights (X_1) are learned. **(c)** (Left) Model performance on the first 400 trials, as compared to the performance reported by [77] (Right)

5.2.4 Discussion

The PFC plays a major role in cognitive control and particularly for learning, selecting and monitoring hierarchical rules. For example, in experimental paradigms, discrimination or categorization tasks can be considered as first-order rules which could be learned individually. However, when conflicting stimuli are presented si-

multaneously, a contextual cue is needed to identify which of the first order rules is to be applied, thus forming second-order rules.

The inner mechanisms of the PFC have been studied in computational models and among them, the property of working memory used for biasing by selective attention in the PBWM model and, more recently in the HER model, the separation between outcome prediction error monitoring, and hierarchical rule learning. Considering the indisputable progress brought by the design of the HER model, we questioned whether it was now a standalone model of the PFC to be used in any circumstances or if the contribution of certain mechanisms like selective attention was still to be considered in some cases and possibly added to the general framework of PFC modeling. More specifically, considering the deployment of cognitive control in realistic behavioral tasks and considering that most hierarchical representations arise from the intersection between agents and the problems they face, and are created over time in a learning process, in a rapid and flexible way, our question was to know if the HER model could account for this kind of process.

Using a task elaborated along two paradigms, we show that, when concrete rules are already learnt and need to be contextualized, the use of a biasing selective attention mechanism is more effective than modulated weights changes in displaying effective cognitive control. When concrete rules are acquired first, superior layers must learn to select the appropriate concrete rule by targeted attention, rather than by relearning representations. We observe that a subject can perform optimally on a given task even though she uses a different rule representation compared to the *official* one. On a single task, this has no consequence and there is actually no way to know which exact rule is used internally. However, when this rule needs to be composed with another rule such as to form a new rule, this may pose problem and lead to bad performance. This has been illustrated on the task: if a subject uses any of the alternative rules for tasks T1 or T2, she'll be unable to solve task T3 even though this task is merely made of a mix of T1 or T2 trials. The reason for the failure of the HER model in this case is to be found in the failure to attend the relevant dimension of the task, here, color, thus claiming for considering and incorporating this mechanism to a versatile PFC model. Analyzing these results in a more general view, we can remark that most experimental paradigms that study hierarchy break down the complexity of a task by providing instructional cues to the participant. Even in studies with rodents and non-human primates, shaping is used in learning paradigms to enable the learning of complex or abstract rules. In developmental learning, this kind of shaping is called curriculum learning. It is evident that such breaking down of complexity must facilitate the acquisition of

abstract rules, and hence modeling approaches must demonstrate these behavioral results.

It is thus important for a model of the PFC to exploit both views, suggesting to incorporate an attentional mechanism for the flexible and controlled design of hierarchical rules from previously learned concrete rules, as we proposed in the new model sketched here.

5.3 Discussion and Perspectives

In this Chapter, we have described a new integrative computational model of cognitive control in the PFC, delineating the roles of (i) mPFC and dlPFC, (ii) posterior and anterior PFC, and (iii) nested CBG loops, building on previous models of hierarchical representations and learning.

Mapping the model to brain regions : In the model, the division of labor between representing abstract rules, or in other words maintaining predictions of possible events versus monitoring errors is assigned to dlPFC and mPFC respectively. Similarly, the distinction between learned concrete rules, and implementing abstract rules, is assigned on the posterior-anterior axis, with the premotor cortex responsible for the former, and more anterior regions (prePMd, mid-dlPFC) for the latter. Further, we make a clear distinction between learning cortical associations in dorsal-medial PFC, versus learning to select the relevant representations through gating with the CBG loops.

Our model proposes a number of mechanisms regarding the interaction of different brain regions. In the HER model, error signals generated by the mPFC are used to train weights that govern how and when a stimulus may be stored in working memory. This is significantly different from the proposed role for BG in WM updating. Consequently, the more plausible reinforcement learning mechanism that we propose is more consistent with the second account, and with the known anatomy of CBG loops. As in the HER model, the medial PFC provides prediction error signals that train the lateral PFC regarding potential errors associated with different if-then rules. The architectural differences included in our model, as borrowed from the theory of nested cortico-striatal loops, where selected abstract rule representations in rostral regions, constrain the selection of concrete rules in caudal regions (through output gating) is a significant difference that allows the system to build on previously acquired information, and thus allows the possibility to generalize learning over multiple contexts (a limitation of the originally proposed HER model). Moreover, the gating mechanism in our model does not restrict the dlPFC's role to maintaining representations over extended periods

of time, but also allows transient activity in the dlPFC to influence rule selection at lower layers.

Thus, the model synthesizes a number of ideas that have previously appeared in the neuroscientific and modeling literature. In light of the evidence we have provided in this chapter, our results confirm that a hierarchically organized system allows individuals to form abstractions that can be applied in a variety of different contexts. Moreover, once discovered, these abstractions or chunks can be applied to novel situations, and help in generalization or transfer learning. Our model provides an architecture that substantiates the intuition that breaking apart a complex task into a multitude of smaller ones will automatically establish a form of hierarchical structure. In the particular experiments we simulate, a set of base competencies is provided (concrete rules, or otherwise policies in RL) upon which more complex learning is based. In this context, our results suggest a bidirectional interaction between attention and learning : attention constrains learning to relevant dimensions of the environment, while the model learns what to attend to via trial and error.

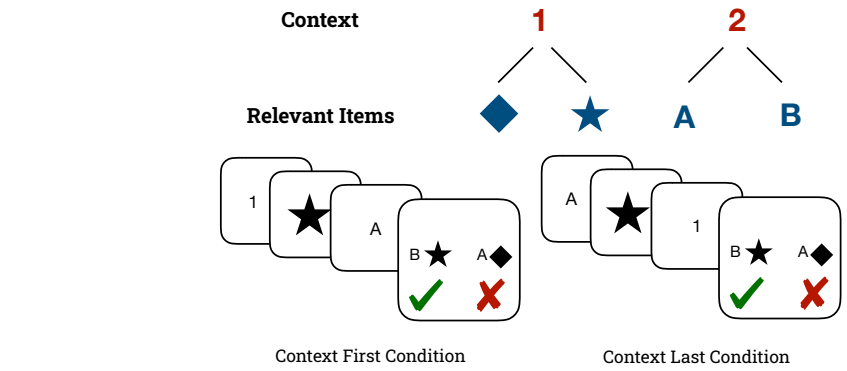
We also highlight that the ability to selectively update some contents of WM while leaving others intact is a process that might be fundamental to hierarchical behavior, because the nature of task representations may have, by definition, different temporal dynamics. Larger goals are relevant over longer periods of time, and thus should not be updated once one sub goal is completed and another is begun.

Relationship to existing theories

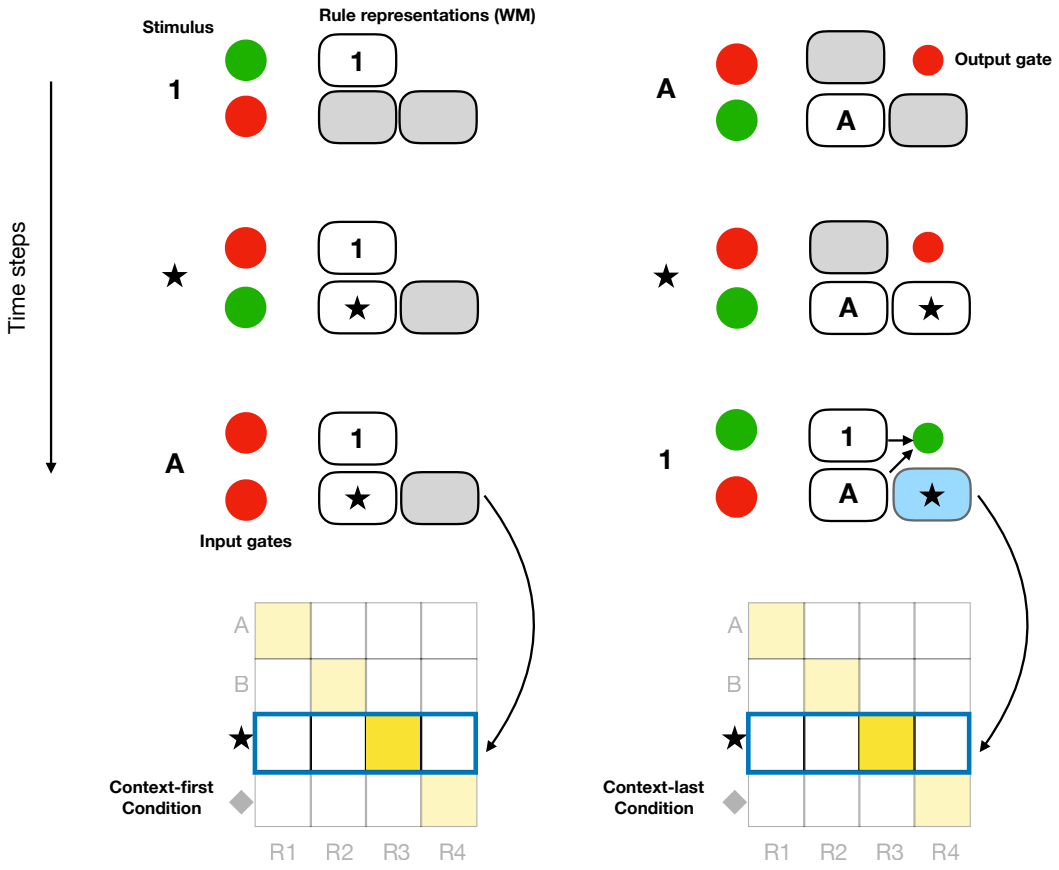
Neurophysiology : To extent of finding the particular brain systems responsible for deploying selective attention, there remains some discrepancy. In the fMRI study by Nee and Brown [145], on the 12AX CPT task described and modeled in previous sections, the authors observed that updating the WM with higher level contextual information involved the anterior PFC and BG. By contrast, updating WM with lower level contextual information involved the posterior PFC and the posterior parietal cortex (PPC). Based on these findings, the authors suggest that responses to lower level context updates may be better modeled as attention shifts than changes in the WM content. Our proposal in the model presented in this chapter is consistent with this account that the presentation of contexts serves to limit the set of relevant rules and engages selective attention towards the currently relevant rules. Interestingly, regions involved in attention are located just anterior to those involved in motor preparation, and consequently the results from this

study could reflect a mixture of motor preparatory processes caudally and attention process rostrally. This could also be due to the fact that at lower levels, the function has already been overtrained and consequently there is no observable activation of the caudal CBG loop, but rather the habitual execution of behavior [163].

Hierarchical organization : Reynolds, O'Reilly, Cohen, and Braver [169] provide a third possible account of the hierarchical organization of the lateral cortex (*adaptive context maintenance hypothesis*), as a middle ground or compromise between the *information cascade hypothesis* and the *levels of abstraction hypothesis*. This theory postulates that both the posterior and anterior regions of PFC are reliably engaged in task conditions requiring active maintenance of contextual information, and that the temporal dynamics of activity in these regions flexibly track the duration of task demands. Areas of both mid-dlPFC and posterior PFC displayed sustained activity when information had to be maintained across multiple trials and more transient activity when information had to be updated frequently.



(a)



(b)

Figure 5.12: **(a)** Illustration of the context-first context-last hierarchical task, as described in [41]. Subjects are presented with a sequence of 3 items - a number, a symbol, and a letter. The numbers represent the context, based on which the relevant items to attend to are either the symbols, or the letters. CF trials are those in which the context appears first, and CL trials are those in which the context appears last. **(b)** Schematic figure explaining how our hierarchical model with selective attention would solve the context last task. Without an output gate to select from within the WM items, the model would be unable to select between the two maintained items in WM. The circles represent the BG input gates, with red signifying a closed gate, and green an open gate. The grids represent the stimulus-action contingencies learned in the lower layer. The task of the higher layers then is to select the appropriate mapping, based on the context.

Experimental Paradigm : To test the demands on input and output gating, Chatham, Frank, and Badre [41] designed an experimental protocol based on hierarchical control tasks requiring the use of conditional rules. In the task, subjects had to base their responses on one of two possible letters, or one of two possible symbols, depending on the identity of a number cue. Thus the number acted as the higher-order context that specified which set of lower order items (letters, or symbols) would be relevant for the response at that trial. This can be seen as a variation on the temporal 12AX task, and the hierarchical categorization tasks described previously, with abstraction at both the temporal, and policy levels. Of note was that each stimulus was presented in an unpredictable serial order on each trial, with the higher order context being presented either before (context first CF) or after (context last CL) the lower order contexts had appeared. Thus, the CF conditions would allow the subjects to use an *input gating* strategy while conversely, under the CL conditions, subjects would need to input each lower order item into the WM and then on the presentation of the context in the last position, select from among the items maintained in the WM that which would influence the response. Thus, the CL condition required the subjects to use an *output gating* strategy. The fMRI results of the experiment showed that activation was greater for contexts presented last relative to first in the lateral frontal and partietal cortices, peaking in prePMd. The posterior parietal cortex was also more strongly recruited for CL than for CF conditions. Consistent with the results from Nee and Brown, the authors failed to observe corticostriatal connectivity during input gating of the lower level context. Instead, corticostriatal coupling increased during the output gating of lower level items. Figure 5.12 illustrates a schematic description of how and when input and output gating acts to solve this task in the two conditions.

Computational modeling : We have already discussed the ideas borrowed in our model from the HER and hierarchical PBWM models. In a different account, the PROBE model [51], models abstract rules as task sets, with each task set represents one strategy stored in long term memory, comprising of

1. a *selective* mapping $Q(s, a)$ which encodes the stimulus-response associations
2. a *predictive* mapping $\gamma(o, s, a)$ which encodes the expected action outcomes given stimuli
3. a *contextual* mapping $F(i|C)$ which encodes external cues predicting the task set reliability

The selection happens at the level of the task set first, then at the level of actions within task sets. Analogous to these computations, in our model, the prediction

weight matrix W at each layer can be thought of as a predictive mapping, and the action values derived from these mappings can be thought of as a selective mapping. However, while the PROBE and C-TS model [53] use RL formulations for learning concrete rules in conjunction with bayesian approaches for updating beliefs about task sets, the model presented here uses the same mechanism to compute error and drive learning.

Another prevalent theory states that quantities related to prediction and calculation of value might constitute the common currency under which the functions of brain regions must be interpreted. The open question then is whether predictive coding in general and the HER model in particular might be expanded to account for the function of additional regions of the PFC without reference to explicit value signalling .

In chapter 2, we have discussed the key findings from neuroscience about the functional subdivisions of the PFC. In this chapter, we focused on the question - "How might the computational modeling of Cognitive Control be influenced by the architectural and algorithmic constraints imposed by the neurobiology of the brain ?" We have shown through our work in this chapter, that a general principle of predictive coding, organized hierarchically is consistent with the neuroscience findings, and that such hierarchy aids in, and may be necessary for the elaboration of contextual or abstract rules. The same general algorithms, with different interpretations of the input and output conditions can explain a lot of the internal information processing likely to be implemented in the brain. In particular, we have highlighted the role of selective attention in working memory, and how this architectural condition can aid in the selection of lower order rules from a possibly large repertoire of learned behaviors.

Conclusion

We began this manuscript with an apparently simple question : How does the brain choose effectively and adaptively among available options to ensure coherent, goal directed behavior ?

The emerging answers to this question from a range of theoretical perspectives have homed in on a consistent set of key computational principles, some of which are shared across species, that emphasize the shaping of behavior at different levels of organization. These principles are implemented in a core set of neural structures that support the valuation, comparison, and selection of behavioral options. The manuscript presented here highlights the achievements and missing elements in the study of cognitive control and proposes a systemic description of the gradient of cognitive control. We described a three fold criteria as the objective of this thesis:

1. to highlight the need for recruiting different mechanisms of cognitive control (to deal with catastrophic forgetting, dealing with the stability-plasticity dilemma and flexibly switching between rules) under different tasks and conditions

2. to algorithmically describe all the mechanisms that fall under the purview of cognitive control in both rodents and primates (working memory, attention, prediction of errors)

3. to model the aforementioned algorithms in a biologically plausible manner (hierarchical organization along the posterior-anterior axis, rule representation versus monitoring along the lateral-medial axis, working memory maintenance and selective attention through CBG loops).

Here, we summarize the contributions of this work, with regards to successfully meeting the criteria identified as missing elements.

In chapter 4, we use the actor critic framework to describe the interaction of multiple learning systems in rodent models, and determine the way that slight differences in task demands can change the demands on these different subsystems in the brains of rodents. We describe how in the simplest case, as permitted by the agranular frontal areas, an immediate decision is sufficient to trigger the behavior

in a stable world, from the selective and predictive models. We hypothesised that mice do possess a "working memory" like capacity, that lets them express contextual behavior. We highlight an oft ignored aspect of cognitive control in computational models - the presence of an innate or "default" behavior i.e. the behavior or actions that occur automatically in response to certain stimuli or contexts. We show that this behavior is not always optimal or desirable, and models need to account for the necessity to override these behaviors in such cases. Our analysis from the three studies is in line with experimental findings from rodent literature, which implicate the dorsal striatum in learning S-R associations and action selection, the mPFC acts as the cognitive control center, implementing a working memory like system, which along with the episodic memory from the hippocampus helps in contextualizing learning. Particularly, in this chapter, we show the limits of cognitive control as seen in rodents, due to the organization and complexity of their brains.

In chapter 5, we show how as permitted by the granular frontal areas, sensory representations must be modified by cognitive control to be adapted to the situation, after an internal deliberation possibly exploiting episodic and semantic memory. We use the HER model as the basis of our computational framework, which suggests that error calculation and representation serve as the common code underlying neural activity and communication. In our study, we investigate the function of specific cerebral structures. Our analysis suggests that the basal ganglia (BG) should be seen as a modulatory system that provides the frontal cortex with flexible gating signals, rather than as a procedural learning system that directly encodes stimulus-response associations, which is the prevailing idea. The evidence supporting this comes from the observation that BG lesions have a greater effect on learning than on behavioral performance. Further, we show how selective attention is deployed by the PFC to guide rule selection at lower levels. We describe not only the interactions between the frontal cortex and the basal ganglia, but also the interactions between the medial and lateral part of the PFC.

However, there is a risk in computational neuroscience of constructing models of specific neural structures in isolation, without considering the broader information flows. In fact, the fundamental biological mechanisms underlying cognitive control have been implemented in various models, but only in isolation and not in an integrative manner. This fractionation of interpretations by specialized subfields may result in an incomplete understanding of the neural mechanisms underlying human behavior. Our description of cognitive control is thus in width rather than depth, and hence evokes a variety of sensorimotor loops and levels of repre-

sentation, from pavlovian to instrumental conditioning, from goal directed to habitual behavior, form episodic to semantic memory, from simple to complex rules, that can coexist and act in competition or in synergy. We propose steps towards integrating key methods and mechanisms to overcome the challenge of developing a unified cognitive architecture.

There is a growing trend in computational neuroscience to develop new models that aim to explain neural processes from scratch, rather than building on existing models and refining them. While this approach can lead to novel insights, it can also result in a proliferation of redundant and overlapping models, which can be difficult to integrate, reconcile and compare. A potential problem with reinventing the wheel is that it can lead to a lack of theoretical coherence, hence diluting the explainability of these models, and eventually making it a challenge to develop a comprehensive understanding of the brain and its functions. We chose in this work, to build on existing models to be able to first, leverage the insights gained from previous research and to be able to address our specific research questions and hypotheses by refining on them. Further, by refining and improving on these models, we are better able to contribute to the ongoing development to the field of cognitive control.

Our presented model of Cognitive Control is by no means complete. For example, uncertainty is defined both as stochasticity (i.e expected uncertainty) and volatility (i.e unexpected uncertainty), and while we give plausible bayesian inference methods to deal with volatility, our model and analysis is limited to deterministic tasks, in which S-R contingencies are strongly defined. In tasks like the multi arm bandit task, in which the reward contingencies behind different actions is probabilistic, and especially in cases in which this probability itself might gradually change over time, it is unclear how our model would be able to deal with such scenarios. A meta learning mechanism may be needed. For example, in the paper by Wang et al [205], the authors show how their RNN is able to tune metaparameters such that after sufficient training, the model "learns how to learn" i.e., it is able to converge on the optimal strategy for the bandit task fairly quickly.

There is evidence of how neurotransmitters like acetylcholine and norepinephrine may play a role in signalling these uncertainty computations [7] [6], and these mechanisms may be integrated into our proposed model at a later stage. Further, we also don't delve into the particular roles of tonic and phasic dopamine in the prediction of errors. There may also be a lateralization effect in the networks associated with cognitive control [85], an avenue which was beyond the scope of this thesis.

Second, it is unclear how and when humans decide to create new rules (or task sets) and hence this problem remains untackled in the model as well. At what point do we decide that none of previously learned knowledge is useful for the task at hand and we must learn a new strategy ?

Perspectives

On the machine learning side, in recent years the emphasis has been on Deep Networks, which are powerful for executing individual tasks, but lack flexibility in several other aspects. State of the art methods remain inferior to human learners in their ability to transfer knowledge to new domains, to capture compositional or systematic structure, to play efficiently and to reason abstractly. The need for using PFC-inspired principles to make deep learning architectures more adaptable to realistic learning times has been identified [179] and in fact the use of attention has become an increasingly popular approach in many tasks. Although the Neural Turing Machines and Long Short-Term Memory architectures have been suggested for the processing of structured and temporal data, they still rely on the same slow and data-intensive learning principles using differentiable functions. Moreover, traditional deep reinforcement learning is slow and incompatible with biological system observations in similar tasks. As a result, episodic reinforcement learning and meta reinforcement learning have been defined, taking inspiration from the hippocampus and prefrontal cortex respectively. However, what is still lacking and could benefit from the current framework is how both learning techniques are linked and interact in their development.

Learned rules are continually refined and updated through experience to become ever more effective in guiding adaptive behavior. In fact, one could argue that the primary job of learning is to extend and enrich internal causal models of the world. This kind of continuous lifelong, online learning can be useful in the context of AI systems that need to learn in real time and adapt their knowledge and skills based on new experiences and data. *Compositionality* and *learning to learn* might be ingredients that make this type of rapid learning possible. When humans or machines make inferences that go beyond the available data, it is likely that strong prior knowledge or inductive biases are compensating for the missing information. *Learning to learn* is a term used to describe how people acquire prior knowledge through previous learning experiences related to a particular task or concept, closely related to the machine learning notions of *transfer learning*, *multi-task learning* or *representational learning*. This prior knowledge can accelerate the learning process of a new task or concept. Hierarchical bayesian model-

ing involves sharing a general prior on concepts among multiple specific concepts, which is learned over time as specific concepts are learned. This approach has been used to explain human learning in various cognitive domains. In machine vision, deep convolutional networks or other discriminative methods can learn-to-learn by sharing features between old and new objects or tasks. Neural networks can also optimize their hyperparameters over a set of related tasks.

Another interesting research perspective is three stages of the development of cognitive control as proposed by Munakata, Snyder, and Chatham [141]. First, children develop the ability to break away from habits and engage in cognitive control in response to environmental stimuli as they grow. The development is linked to their ability to retain information relevant to their goals, starting with concrete information, such as remembering the location of a hidden toy, and gradually moving towards more abstract information, like task rules. These actively maintained representations help support their goal directed thoughts and behaviors. As they continue to develop, children start recruiting cognitive control more proactively, anticipating the need for it instead of reacting to the moment. They also become more self directed in their cognitive control, relying less on external signals and more on their internal representations.

A complete picture of cognitive control is also incomplete without reference to the interactions between the PFC and the hippocampus. Evidence suggests that abstracted representations in the medial prefrontal cortex (mPFC) guide the reactivation of related memories during new encoding events, which promotes the integration of related experiences in the hippocampus. Recent studies have also shown that integrated memories are used during novel situations to facilitate a range of behaviors such as spatial navigation and imagination [183]. Moreover, information that is congruent with existing knowledge (a schema) is usually better remembered than less congruent information and thus, another learning system may be needed to overcome interference resulting from multiple medial temporal lobe (MTL) instances sharing common elements. The function of the mPFC may then be to select the most relevant elements of an experience during both encoding and retrieval, and suppress activity in representations inconsistent with the dominant schema while amplifying activity in congruent representations [200]. The ability to imagine hypothetical situations before they happen can increase the accuracy of predicting their outcome. Remembering past experiences is a reconstructive process where memories are recreated from their component parts. Construction may therefore crucial for planning the future and recalling the past [90].

Bibliography

1. G. E. Alexander, M. R. DeLong, and P. L. Strick. "Parallel organization of functionally segregated circuits linking basal ganglia and cortex". *Annual review of neuroscience* 9:1, 1986, pp. 357–381.
2. W. H. Alexander and J. W. Brown. "Frontal cortex function as derived from hierarchical predictive coding". *Scientific reports* 8:1, 2018, p. 3843.
3. W. H. Alexander and J. W. Brown. "Hierarchical error representation: a computational model of anterior cingulate and dorsolateral prefrontal cortex". *Neural Computation* 27:11, 2015, pp. 2354–2410.
4. W. H. Alexander and J. W. Brown. "Medial prefrontal cortex as an action-outcome predictor". *Nature neuroscience* 14:10, 2011, pp. 1338–1344.
5. F. Alexandre. "A global framework for a systemic view of brain modeling". *Brain Informatics* 8:1, 2021, p. 3.
6. F. Alexandre and M. Carrere. "Modeling neuromodulation as a framework to integrate uncertainty in general cognitive architectures". In: *Artificial General Intelligence: 9th International Conference, AGI 2016, New York, NY, USA, July 16-19, 2016, Proceedings 9*. Springer. 2016, pp. 324–333.
7. J. Y. Angela and P. Dayan. "Uncertainty, neuromodulation, and attention". *Neuron* 46:4, 2005, pp. 681–692.
8. E. Antonelo and B. Schrauwen. "Learning slow features with reservoir computing for biologically-inspired robot localization". *Neural Networks* 25, 2012, pp. 178–190.
9. E. A. Antonelo and B. Schrauwen. "On learning navigation behaviors for small mobile robots with reservoir computing architectures". *IEEE transactions on neural networks and learning systems* 26:4, 2014, pp. 763–780.
10. W. F. Asaad, G. Rainer, and E. K. Miller. "Task-specific neural activity in the primate prefrontal cortex". *Journal of neurophysiology* 84:1, 2000, pp. 451–459.
11. D. Badre and M. D'esposito. "Is the rostro-caudal axis of the frontal lobe hierarchical?" *Nature Reviews Neuroscience* 10:9, 2009, pp. 659–669.

12. D. Badre, J. Hoffman, J. W. Cooney, and M. D'esposito. "Hierarchical cognitive control deficits following damage to the human frontal lobe". *Nature neuroscience* 12:4, 2009, pp. 515–522.

13. D. Badre, A. S. Kayser, and M. D'Esposito. "Frontal Cortex and the Discovery of Abstract Action Rules". *Neuron* 66:2, 2010, pp. 315–326. DOI: 10.1016/j.neuron.2010.03.025. URL: https://doi.org/10.1016%2Fj.neuron.2010.03.025.

14. D. Badre and D. E. Nee. "Frontal cortex and the hierarchical control of behavior". *Trends in cognitive sciences* 22:2, 2018, pp. 170–188.

15. J. Bahlmann, R. S. Blumenfeld, and M. D'Esposito. "The rostro-caudal axis of frontal cortex is sensitive to the domain of stimulus information". *Cerebral cortex* 25:7, 2015, pp. 1815–1826.

16. L. W. Barsalou. "Frames, concepts, and conceptual fields.", 1992.

17. O. Bartra, J. T. McGuire, and J. W. Kable. "The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value". *Neuroimage* 76, 2013, pp. 412–427.

18. M. G. Baxter, A. Parker, C. C. Lindner, A. D. Izquierdo, and E. A. Murray. "Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex". *Journal of Neuroscience* 20:11, 2000, pp. 4311–4319.

19. T. E. Behrens, M. W. Woolrich, M. E. Walton, and M. F. Rushworth. "Learning the value of information in an uncertain world". *Nature neuroscience* 10:9, 2007, pp. 1214–1221.

20. J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, and D. D. Cox. "Hyperopt: a python library for model selection and hyperparameter optimization". *Computational Science & Discovery* 8:1, 2015, p. 014008.

21. A. Bhandari and D. Badre. "Learning and transfer of working memory gating policies". *Cognition* 172, 2018, pp. 89–100.

22. E. D. Boorman, T. E. Behrens, M. W. Woolrich, and M. F. Rushworth. "How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action". *Neuron* 62:5, 2009, pp. 733–743.

23. E. D. Boorman and M. Noonan. "Contributions of ventromedial prefrontal and frontal polar cortex to reinforcement learning and value-based choice". *Neural basis of motivational and cognitive control*, 2011, pp. 55–74.

24. M. Botvinick and T. Braver. "Motivation and cognitive control: from behavior to neural mechanism". *Annual review of psychology* 66, 2015, pp. 83–113.

25. M. Botvinick, S. Ritter, J. X. Wang, Z. Kurth-Nelson, C. Blundell, and D. Hassabis. "Reinforcement learning, fast and slow". *Trends in cognitive sciences* 23:5, 2019, pp. 408–422.

26. M. M. Botvinick. "Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function". *Cognitive, Affective, & Behavioral Neuroscience* 7:4, 2007, pp. 356–366.

27. M. M. Botvinick, T. S. Braver, D. M. Barch, C. S. Carter, and J. D. Cohen. "Conflict monitoring and cognitive control." *Psychological review* 108:3, 2001, p. 624.

28. M. M. Botvinick, S. Huffstetler, and J. T. McGuire. "Effort discounting in human nucleus accumbens". *Cognitive, affective, & behavioral neuroscience* 9:1, 2009, pp. 16–27.

29. F. Bouchacourt, S. Palminteri, E. Koechlin, and S. Ostojic. "Temporal chunking as a mechanism for unsupervised learning of task-sets". *Elife* 9, 2020, e50469.

30. T. S. Braver. "The variable nature of cognitive control: a dual mechanisms framework". *Trends in cognitive sciences* 16:2, 2012, pp. 106–113.

31. N. J. Broadbent, L. R. Squire, and R. E. Clark. "Rats depend on habit memory for discrimination learning and retention". *Learning & Memory* 14:3, 2007, pp. 145–151.

32. J. W. Brown and T. S. Braver. "Learned predictions of error likelihood in the anterior cingulate cortex". *Science* 307:5712, 2005, pp. 1118–1121.

33. S. A. Bunge. "How we use rules to select actions: a review of evidence from cognitive neuroscience". *Cognitive, Affective, & Behavioral Neuroscience* 4:4, 2004, pp. 564–579.

34. S. A. Bunge, I. Kahn, J. D. Wallis, E. K. Miller, and A. D. Wagner. "Neural circuits subserving the retrieval and maintenance of abstract rules". *Journal of neurophysiology* 90:5, 2003, pp. 3419–3428.

35. S. A. Bunge, K. N. Ochsner, J. E. Desmond, G. H. Glover, and J. D. Gabrieli. "Prefrontal regions involved in keeping information in and out of mind". *Brain* 124:10, 2001, pp. 2074–2086.

36. T. J. Bussey, J. L. Muir, B. J. Everitt, and T. W. Robbins. "Dissociable effects of anterior and posterior cingulate cortex lesions on the acquisition of a conditional visual discrimination: facilitation of early learning vs. impairment of late learning". *Behavioural brain research* 82:1, 1996, pp. 45–56.

37. T. J. Bussey, J. L. Muir, B. J. Everitt, and T. W. Robbins. "Triple dissociation of anterior cingulate, posterior cingulate, and medial frontal cortices on visual discrimination tasks using a touchscreen testing procedure for the rat." *Behavioral neuroscience* 111:5, 1997, p. 920.

38. T. J. Bussey, S. P. Wise, and E. A. Murray. "The role of ventral and orbital prefrontal cortex in conditional visuomotor learning and strategy use in rhesus monkeys (Macaca mulatta)." *Behavioral neuroscience* 115:5, 2001, p. 971.

39. N. Chaix-Eichel, S. Dagar, Q. Lanneau, K. Sobriel, T. Boraud, F. Alexandre, and N. P. Rougier. "From implicit learning to explicit representations". *arXiv preprint arXiv:2204.02484*, 2022.

40. J.-P. Changeux and S. Dehaene. "Neuronal models of cognitive functions". *Cognition* 33:1-2, 1989, pp. 63–109.

41. C. H. Chatham, M. J. Frank, and D. Badre. "Corticostriatal output gating during selection from working memory". *Neuron* 81:4, 2014, pp. 930–942.

42. P. A. Chouinard and T. Paus. "The primary motor and premotor areas of the human cerebral cortex". *The neuroscientist* 12:2, 2006, pp. 143–152.

43. Y. Chudasama and T. W. Robbins. "Dissociable contributions of the orbitofrontal and infralimbic cortex to pavlovian autoshaping and discrimination reversal learning: further evidence for the functional heterogeneity of the rodent frontal cortex". *Journal of Neuroscience* 23:25, 2003, pp. 8771–8780.

44. N. Clairis and M. Pessiglione. "Value, confidence, deliberation: a functional partition of the medial prefrontal cortex demonstrated across rating and choice tasks". *Journal of Neuroscience*, 2022.

45. A. Clark and A. Karmiloff-Smith. "The cognizer's innards: A philosophical and psychological perspective on the development of thought", 1993.

46. H. F. Clarke, T. W. Robbins, and A. C. Roberts. "Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex". *Journal of Neuroscience* 28:43, 2008, pp. 10972–10982.

47. A. Cleeremans. "Implicit learning and implicit memory", 2009.

48. A. Cleeremans, B. Timmermans, and A. Pasquali. "Consciousness and metarepresentation: A computational sketch". *Neural Networks* 20:9, 2007, pp. 1032–1039.

49. B. A. Clegg, G. J. DiGirolamo, and S. W. Keele. "Sequence learning". *Trends in cognitive sciences* 2:8, 1998, pp. 275–281.

50. J. D. Cohen, T. S. Braver, and R. O'Reilly. "A computational approach to prefrontal cortex, cognitive control and schizophrenia: recent developments and current challenges". *Philosophical transactions of the royal society of london. Series B: Biological sciences* 351:1346, 1996, pp. 1515–1527.

51. A. Collins and E. Koechlin. "Reasoning, learning, and creativity: frontal lobe function and human decision-making". *PLoS biology* 10:3, 2012, e1001293.

52. A. G. E. Collins and J. Cockburn. "Beyond dichotomies in reinforcement learning". *Nature Reviews Neuroscience* 21:10, 2020, pp. 576–586. DOI: 10.1038/s41583-020-0355-6. URL: https://doi.org/10.1038%2Fs41583-020-0355-6.

53. A. G. Collins and M. J. Frank. "Cognitive control over learning: creating, clustering, and generalizing task-set structure." *Psychological review* 120:1, 2013, p. 190.

54. E. Coutureau and S. Killcross. "Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats". *Behavioural Brain Research* 146:1, 2003. The Rodent Prefrontal Cortex, pp. 167–174. ISSN: 0166-4328. DOI: https://doi.org/10.1016/j.bbr.2003.09.025. URL: https://www.sciencedirect.com/science/article/pii/S0166432803003498.

55. B. C. Da Silva, E. W. Basso, A. L. Bazzan, and P. M. Engel. "Dealing with non-stationary environments using context detection". In: *Proceedings of the 23rd international conference on Machine learning*. 2006, pp. 217–224.

56. S. Dagar, F. Alexandre, and N. Rougier. "Deciphering the contributions of episodic and working memories in increasingly complex decision tasks". In: *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2021, pp. 1–6.

57. N. D. Daw, Y. Niv, and P. Dayan. "Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control". *Nature Neuroscience* 8:12, 2005, pp. 1704–1711. DOI: 10.1038/nn1560. URL: https://doi.org/10.1038%2Fnn1560.

58. P. Dayan and B. Seymour. "Values and actions in aversion". In: *Neuroeconomics*. Elsevier, 2009, pp. 175–191.

59. P. Dayan and A. J. Yu. "Expected and unexpected uncertainty: ACh and NE in the neocortex". *Advances in neural information processing systems* 15, 2002.

60. G. Deco and E. T. Rolls. "Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex". *European Journal of Neuroscience* 18:8, 2003, pp. 2374–2390.

61. A. S. Dekaban and D. Sadowsky. "Changes in brain weights during the span of human life: relation of brain weights to body heights and body weights". *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 4:4, 1978, pp. 345–356.

62. Z. Dienes and D. Berry. "Implicit learning: Below the subjective threshold". *Psychonomic bulletin & review* 4:1, 1997, pp. 3–23.

63. P. Domenech and E. Koechlin. "Executive control and decision-making in the prefrontal cortex". *Current opinion in behavioral sciences* 1, 2015, pp. 101–106.

64. P. Domenech, S. Rheims, and E. Koechlin. "Neural mechanisms resolving exploitation-exploration dilemmas in the medial prefrontal cortex". *Science* 369:6507, 2020, eabb0184.

65. P. Dominey, M. Arbib, and J.-P. Joseph. "A model of corticostriatal plasticity for learning oculomotor associations and sequences". *Journal of cognitive neuroscience* 7:3, 1995, pp. 311–336.

66. N. U. Dosenbach, K. M. Visscher, E. D. Palmer, F. M. Miezin, K. K. Wenger, H. C. Kang, E. D. Burgund, A. L. Grimes, B. L. Schlaggar, and S. E. Petersen. "A core system for the implementation of task sets". *Neuron* 50:5, 2006, pp. 799–812.

67. K. Doya, K. Samejima, K.-i. Katagiri, and M. Kawato. "Multiple model-based reinforcement learning". *Neural computation* 14:6, 2002, pp. 1347–1369.

68. M. K. Eckstein and A. G. Collins. "Computational evidence for hierarchically structured reinforcement learning in humans". *Proceedings of the National Academy of Sciences* 117:47, 2020, pp. 29381–29389.

69. H. Eichenbaum. "Memory: organization and control". *Annual review of psychology* 68, 2017, p. 19.

70. R. W. Engle and M. J. Kane. "Executive attention, working memory capacity, and a two-factor theory of cognitive control.", 2004.

71. +. Faure, +. Haberland, +. Condé, and +. Massioui. "Lesion to the Nigrostriatal Dopamine System Disrupts Stimulus-Response Habit Formation". *J. Neurosci.; Journal of Neuroscience* 25:11, 2005, pp. 2771–2780.

72. R. Featherstone and R. McDonald. "Lesions of the dorsolateral striatum impair the acquisition of a simplified stimulus-response dependent conditional discrimination task". *Neuroscience* 136:2, 2005, pp. 387–395.

73. L. K. Fellows. "The role of orbitofrontal cortex in decision making: a component process account". *Annals of the New York Academy of Sciences* 1121:1, 2007, pp. 421–430.

74. T. H. FitzGerald, B. Seymour, and R. J. Dolan. "The role of human orbitofrontal cortex in value comparison for incommensurable objects". *Journal of Neuroscience* 29:26, 2009, pp. 8388–8395.

75. P. Fleischer and S. Hélie. "A unified model of rule-set learning and selection". *Neural Networks* 124, 2020, pp. 343–356.

76. L. M. Frank, E. N. Brown, and M. Wilson. "Trajectory encoding in the hippocampus and entorhinal cortex". *Neuron* 27:1, 2000, pp. 169–178.

77. M. J. Frank and D. Badre. "Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: computational analysis". *Cerebral cortex* 22:3, 2012, pp. 509–526.

78. D. J. Freedman, M. Riesenhuber, T. Poggio, and E. K. Miller. "Categorical representation of visual stimuli in the primate prefrontal cortex". *Science* 291:5502, 2001, pp. 312–316.

79. J. F. Fulton. "A note on the definition of the "motor" and "premotor" areas". *Brain* 58:2, 1935, pp. 311–316.

80. J. M. Fuster. "The prefrontal cortex—an update: time is of the essence". *Neuron* 30:2, 2001, pp. 319–333.

81. L. J. Garey. *Brodmann's' localisation in the cerebral cortex'*. World Scientific, 1999.

82. A. Genovesio, P. J. Brasted, A. R. Mitz, and S. P. Wise. "Prefrontal cortex activity related to abstract response strategies". *Neuron* 47:2, 2005, pp. 307–320.

83. D. Gentner and J. Medina. "Similarity and the development of rules". *Cognition* 65:2-3, 1998, pp. 263–297.

84. M. Gilead, Y. Trope, and N. Liberman. "Above and beyond the concrete: The diverse representational substrates of the predictive brain". *Behavioral and Brain Sciences* 43, 2020.

85. J. Gläscher, R. Adolphs, H. Damasio, A. Bechara, D. Rudrauf, M. Calamia, L. K. Paul, and D. Tranel. "Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex". *Proceedings of the National Academy of Sciences* 109:36, 2012, pp. 14681–14686.

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

87. S. Grossberg. "Competitive learning: From interactive activation to adaptive resonance". *Cognitive science* 11:1, 1987, pp. 23–63.

88. M. Grueschow, R. Polania, T. A. Hare, and C. C. Ruff. "Automatic versus choice-dependent value representations in the human brain". *Neuron* 85:4, 2015, pp. 874–885.

89. A. Habedank, P. Kahnau, and L. Lewejohann. "Alternate without alternative: neither preference nor learning explains behaviour of C57BL/6J mice in the T-maze". *Behaviour* 158:7, 2021, pp. 625–662.

90. D. Hassabis and E. A. Maguire. "The construction system of the brain". *Philosophical Transactions of the Royal Society B: Biological Sciences* 364:1521, 2009, pp. 1263–1271.

91. B. Y. Hayden and M. L. Platt. "Neurons in anterior cingulate cortex multiplex information about reward and action". *Journal of Neuroscience* 30:9, 2010, pp. 3339–3346.

92. T. E. Hazy, M. J. Frank, and R. C. O'Reilly. *Computational neuroscientific models of working memory*. 2021.

93. S. Herbert et al. "The architecture of complexity". *Proceedings of the American Philosophical Society* 106:6, 1962, pp. 467–482.

94. C. B. Holroyd and M. G. Coles. "Dorsal anterior cingulate cortex integrates reinforcement history to guide voluntary behavior". *cortex* 44:5, 2008, pp. 548–559.

95. C. B. Holroyd and M. G. Coles. "The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity." *Psychological review* 109:4, 2002, p. 679.

96. C. B. Holroyd and S. M. McClure. "Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model." *Psychological review* 122:1, 2015, p. 54.

97. C. B. Holroyd and N. Yeung. "Motivation of extended behaviors by anterior cingulate cortex". *Trends in cognitive sciences* 16:2, 2012, pp. 122–128.

98. E. Hoshi. "Differential involvement of the prefrontal, premotor, and primary motor cortices in rule-based motor behavior". *Neuroscience of rule-guided behavior*, 2008, pp. 159–175.

99. L. T. Hunt, N. Kolling, A. Soltani, M. W. Woolrich, M. F. Rushworth, and T. E. Behrens. "Mechanisms underlying cortical activity during value-guided choice". *Nature neuroscience* 15:3, 2012, pp. 470–476.

100. I. R. Inglis, S. Langton, B. Forkman, and J. Lazarus. "An information primacy model of exploratory and foraging behaviour". *Animal behaviour* 62:3, 2001, pp. 543–557.

101. A. Izquierdo, R. K. Suda, and E. A. Murray. "Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency". *Journal of Neuroscience* 24:34, 2004, pp. 7540–7548.

102. G. Jocham, T. A. Klein, and M. Ullsperger. "Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices". *Journal of Neuroscience* 31:5, 2011, pp. 1606–1613.

103. M. Jones and B. C. Love. "Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition". *Behavioral and brain sciences* 34:4, 2011, pp. 169–188.

104. M. J. Kane, M. K. Bleckley, A. R. Conway, and R. W. Engle. "A controlled-attention view of working-memory capacity." *Journal of experimental psychology: General* 130:2, 2001, p. 169.

105. D. B. Kastner, A. K. Gillespie, P. Dayan, and L. M. Frank. "Memory alone does not account for the way rats learn a simple spatial alternation task". *Journal of Neuroscience* 40:38, 2020, pp. 7311–7317.

106. S. W. Kennerley, T. E. Behrens, and J. D. Wallis. "Double dissociation of value computations in orbitofrontal and anterior cingulate neurons". *Nature neuroscience* 14:12, 2011, pp. 1581–1589.

107. M. Khamassi, S. Lallée, P. Enel, E. Procyk, and P. F. Dominey. "Robot cognitive control with a neurophysiologically inspired reinforcement learning model". *Frontiers in neurorobotics* 5, 2011, p. 1.

108. M. C. Klein-Flügge, A. Bongioanni, and M. F. Rushworth. "Medial and orbital frontal cortex in decision-making and flexible behavior". *Neuron*, 2022.

109. M. C. Klein-Flügge, S. W. Kennerley, K. Friston, and S. Bestmann. "Neural signatures of value comparison in human cingulate cortex during decisions requiring an effort-reward trade-off". *Journal of Neuroscience* 36:39, 2016, pp. 10002–10015.

110. B. Knutson, J. Taylor, M. Kaufman, R. Peterson, and G. Glover. "Distributed neural representation of expected value". *Journal of Neuroscience* 25:19, 2005, pp. 4806–4812.

111. E. Koechlin. "An evolutionary computational theory of prefrontal executive function in decision-making". *Philosophical Transactions of the Royal Society B: Biological Sciences* 369:1655, 2014, p. 20130474.

112. E. Koechlin, C. Ody, and F. Kouneiher. "The architecture of cognitive control in the human prefrontal cortex". *Science* 302:5648, 2003, pp. 1181–1185.

113. E. Koechlin and C. Summerfield. "An information theoretical approach to prefrontal executive function". *Trends in cognitive sciences* 11:6, 2007, pp. 229–235.

114. N. Kolling, T. E. Behrens, R. B. Mars, and M. F. Rushworth. "Neural mechanisms of foraging". *Science* 336:6077, 2012, pp. 95–98.

115. F. Kouneiher, S. Charron, and E. Koechlin. "Motivation and cognitive control in the human prefrontal cortex". *Nature neuroscience* 12:7, 2009, pp. 939–945.

116. P. Krack, M. I. Hariz, C. Baunez, J. Guridi, and J. A. Obeso. "Deep brain stimulation: from neurology to psychiatry?" *Trends in neurosciences* 33:10, 2010, pp. 474–484.

117. J. W. Krakauer, A. A. Ghazanfar, A. Gomez-Marin, M. A. MacIver, and D. Poeppel. "Neuroscience needs behavior: correcting a reductionist bias". *Neuron* 93:3, 2017, pp. 480–490.

118. J. Krettek and J. Price. "The cortical projections of the mediodorsal nucleus and adjacent thalamic nuclei in the rat". *Journal of comparative neurology* 171:2, 1977, pp. 157–191.

119. M. L. Kringelbach. "The human orbitofrontal cortex: linking reward to hedonic experience". *Nature reviews neuroscience* 6:9, 2005, pp. 691–702.

120. W. Kruijne, S. M. Bohte, P. R. Roelfsema, and C. N. Olivers. "Flexible working memory through selective gating and attentional tagging". *Neural Computation* 33:1, 2021, pp. 1–40.

121. B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman. "Building machines that learn and think like people". *Behavioral and brain sciences* 40, 2017.

122. G. P. Latham and E. A. Locke. "Goal setting—A motivational technique that works". *Organizational dynamics* 8:2, 1979, pp. 68–80.

123. M. Laubach, L. M. Amarante, K. Swanson, and S. R. White. "What, if anything, is rodent prefrontal cortex?" *eneuro* 5:5, 2018.

124. D. Lee, H. Seo, and M. W. Jung. "Neural basis of reinforcement learning and decision making". *Annual review of neuroscience* 35, 2012, p. 287.

125. R. K. Lenroot and J. N. Giedd. "Brain development in children and adolescents: insights from anatomical magnetic resonance imaging". *Neuroscience & biobehavioral reviews* 30:6, 2006, pp. 718–729.

126. H. J. Levesque. "On our best behaviour". *Artificial Intelligence* 212, 2014, pp. 27–35.

127. D. J. Levy and P. W. Glimcher. "The root of all value: a neural common currency for choice". *Current opinion in neurobiology* 22:6, 2012, pp. 1027–1038.

128. S.-L. Lim, J. P. O'Doherty, and A. Rangel. "The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention". *Journal of Neuroscience* 31:37, 2011, pp. 13214–13223.

129. K. Lloyd, N. Becker, M. W. Jones, and R. Bogacz. "Learning to use working memory: a reinforcement learning gating model of rule acquisition in rats". *Frontiers in computational neuroscience* 6, 2012, p. 87.

130. K. Lloyd and D. S. Leslie. "Context-dependent decision-making: a simple Bayesian model". *Journal of The Royal Society Interface* 10:82, 2013, p. 20130069.

131. P. Lopez-Gamundi, Y.-W. Yao, T. T. Chong, H. R. Heekeren, E. Mas-Herrero, and J. Marco-Pallarés. "The neural basis of effort valuation: A meta-analysis of functional magnetic resonance imaging studies". *Neuroscience & Biobehavioral Reviews* 131, 2021, pp. 1275–1287.

132. E. Maes, G. De Filippo, A. B. Inkster, S. E. Lea, J. De Houwer, R. D'Hooge, T. Beckers, and A. J. Wills. "Feature-versus rule-based generalization in rats, pigeons and humans". *Animal Cognition* 18:6, 2015, pp. 1267–1284.

133. F. A. Mansouri, D. J. Freedman, and M. J. Buckley. "Emergence of abstract rules in the primate brain". *Nature Reviews Neuroscience* 21:11, 2020, pp. 595–610.

134. D. Marr and T. Poggio. "A computational theory of human stereo vision". *Proceedings of the Royal Society of London. Series B. Biological Sciences* 204:1156, 1979, pp. 301–328.

135. M. Martinolli, W. Gerstner, and A. Gilra. "Multi-Timescale memory dynamics extend task repertoire in a reinforcement learning network with Attention-Gated memory". *Frontiers in computational neuroscience* 12, 2018, p. 50.

136. U. Mayr and R. Kliegl. "Task-set switching and long-term memory retrieval.", 2000.

137. S. M. McClure, D. I. Laibson, G. Loewenstein, and J. D. Cohen. "Separate neural systems value immediate and delayed monetary rewards". *Science* 306:5695, 2004, pp. 503–507.

138. K. J. Miller, E. A. Ludvig, G. Pezzulo, and A. Shenhav. "Realigning models of habitual and goal-directed decision-making". In: *Goal-directed decision making*. Elsevier, 2018, pp. 407–428.

139. T. Minamimoto, R. C. Saunders, and B. J. Richmond. "Monkeys quickly learn and generalize visual categories without lateral prefrontal cortex". *Neuron* 66:4, 2010, pp. 501–507.

140. N. G. Müller and R. T. Knight. "The functional neuroanatomy of working memory: contributions of human brain lesion studies". *Neuroscience* 139:1, 2006, pp. 51–58.

141. Y. Munakata, H. R. Snyder, and C. H. Chatham. "Developing cognitive control: Three key transitions". *Current directions in psychological science* 21:2, 2012, pp. 71–77.

142. E. A. Murray and P. H. Rudebeck. "Specializations for reward-guided decision-making in the primate ventral prefrontal cortex". *Nature Reviews Neuroscience* 19:7, 2018, pp. 404–417.

143. K. Nakahara, T. Hayashi, S. Konishi, and Y. Miyashita. "Functional MRI of macaque monkeys performing a cognitive set-shifting task". *Science* 295:5559, 2002, pp. 1532–1536.

144. M. R. Nassar, R. C. Wilson, B. Heasly, and J. I. Gold. "An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment". *Journal of Neuroscience* 30:37, 2010, pp. 12366–12378.

145. D. E. Nee and J. W. Brown. "Dissociable frontal–striatal and frontal–parietal networks involved in updating hierarchical contexts in working memory". *Cerebral cortex* 23:9, 2013, pp. 2146–2158.

146. A. Nieder, D. J. Freedman, and E. K. Miller. "Representation of the quantity of visual items in the primate prefrontal cortex". *Science* 297:5587, 2002, pp. 1708–1711.

147. Y. Niv. "The primacy of behavioral research for understanding the brain." *Behavioral Neuroscience* 135:5, 2021, p. 601.

148. Y. Niv, N. D. Daw, D. Joel, and P. Dayan. "Tonic dopamine: opportunity costs and the control of response vigor". *Psychopharmacology* 191, 2007, pp. 507–520.

149. M. Noonan, R. Mars, and M. Rushworth. "Distinct roles of three frontal cortical areas in reward-guided behavior". *Journal of Neuroscience* 31:40, 2011, pp. 14399–14412.

150. R. C. O'Reilly and M. J. Frank. "Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia". *Neural computation* 18:2, 2006, pp. 283–328.

151. R. C. O'Reilly, D. C. Noelle, T. S. Braver, and J. D. Cohen. "Prefrontal cortex and dynamic categorization tasks: representational organization and neuromodulatory control". *Cerebral cortex* 12:3, 2002, pp. 246–257.

152. P.-Y. Oudeyer and F. Kaplan. "What is intrinsic motivation? A typology of computational approaches". *Frontiers in neurorobotics*, 2009, p. 6.

153. M. G. Packard and J. L. McGaugh. "Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning". *Neurobiology of learning and memory* 65:1, 1996, pp. 65–72.

154. C. Padoa-Schioppa. "Neurobiology of economic choice: a good-based model". *Annual review of neuroscience* 34, 2011, pp. 333–359.

155. R. E. Passingham. *The frontal lobes and voluntary action*. Vol. 21. OUP Oxford, 1995.

156. D. C. Penn, K. J. Holyoak, and D. J. Povinelli. "Darwin's mistake: Explaining the discontinuity between human and nonhuman minds". *Behavioral and brain sciences* 31:2, 2008, pp. 109–130.

157. L. Peshkin, N. Meuleau, and L. Kaelbling. "Learning policies with external memory". *arXiv preprint cs/0103003*, 2001.

158. M. Pessiglione, F. Vinckier, S. Bouret, J. Daunizeau, and R. Le Bouc. "Why not try harder? Computational approach to motivation deficits in neuropsychiatric diseases". *Brain* 141:3, 2018, pp. 629–650.

159. L. Pessoa. "The entangled brain". *Journal of Cognitive Neuroscience*, 2022, pp. 1–12.

160. M. Petrides. "Frontal lobes and working memory: evidence from investigations of the effects of cortical excisions in nonhuman primates". *Handbook of neuropsychology* 9, 1994, pp. 59–82.

161. M. Petrides. "Impairments on nonspatial self-ordered and externally ordered working memory tasks after lesions of the mid-dorsal part of the lateral frontal cortex in the monkey". *Journal of Neuroscience* 15:1, 1995, pp. 359–375.

162. N. Picard and P. L. Strick. "Imaging the premotor areas". *Current opinion in neurobiology* 11:6, 2001, pp. 663–672.

163. C. Piron, D. Kase, M. Topalidou, M. Goillandeau, H. Orignac, T.-H. N'Guyen, N. Rougier, and T. Boraud. "The globus pallidus pars interna in goal-oriented and routine behaviors: resolving a long-standing paradox". *Movement disorders* 31:8, 2016, pp. 1146–1154.

164. E. M. Pothos. "The rules versus similarity distinction". *Behavioral and brain sciences* 28:1, 2005, pp. 1–14.

165. T. M. Preuss and S. P. Wise. "Evolution of prefrontal cortex". *Neuropsychopharmacology* 47:1, 2022, pp. 3–19.

166. D. Purves, R. Cabeza, S. A. Huettel, K. S. LaBar, M. L. Platt, M. G. Woldorff, and E. M. Brannon. *Cognitive neuroscience*. Vol. 6. 4. Sunderland: Sinauer Associates, Inc, 2008.

167. A. S. Reber. "Implicit learning and tacit knowledge." *Journal of experimental psychology: General* 118:3, 1989, p. 219.

168. R. A. Rescorla. "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement". *Classical conditioning, Current research and theory* 2, 1972, pp. 64–69.

169. J. R. Reynolds, R. C. O'Reilly, J. D. Cohen, and T. S. Braver. "The function and organization of lateral prefrontal cortex: a test of competing hypotheses". *PloS one* 7:2, 2012, e30284.

170. J. J. Ribas-Fernandes, A. Solway, C. Diuk, J. T. McGuire, A. G. Barto, Y. Niv, and M. M. Botvinick. "A neural signature of hierarchical reinforcement learning". *Neuron* 71:2, 2011, pp. 370–379.

171. S. Ritter, J. X. Wang, Z. Kurth-Nelson, and M. Botvinick. "Episodic control as meta-reinforcement learning". *BioRxiv*, 2018, p. 360537.

172. J. O. Rombouts, S. M. Bohte, and P. R. Roelfsema. "How attention can create synaptic tags for the learning of working memories in sequential tasks". *PLoS computational biology* 11:3, 2015, e1004060.

173. N. Rougier, C. Brun, and T. Boraud. "A Mathematical Bias", 2021.

174. J. S. Rubinstein, D. E. Meyer, and J. E. Evans. "Executive control of cognitive processes in task switching." *Journal of experimental psychology: human perception and performance* 27:4, 2001, p. 763.

175. P. H. Rudebeck, T. E. Behrens, S. W. Kennerley, M. G. Baxter, M. J. Buckley, M. E. Walton, and M. F. Rushworth. "Frontal cortex subregions play distinct roles in choices between actions and stimuli". *Journal of Neuroscience* 28:51, 2008, pp. 13775–13785.

176. M. F. Rushworth, T. Behrens, P. Rudebeck, and M. Walton. "Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour". *Trends in cognitive sciences* 11:4, 2007, pp. 168–176.

177. M. F. Rushworth, M. J. Buckley, T. E. Behrens, M. E. Walton, and D. M. Bannerman. "Functional organization of the medial frontal cortex". *Current opinion in neurobiology* 17:2, 2007, pp. 220–227.

178. M. Rushworth, M. E. Walton, S. W. Kennerley, and D. Bannerman. "Action sets and decisions in the medial frontal cortex". *Trends in cognitive sciences* 8:9, 2004, pp. 410–417.

179. J. Russin, R. C. O'Reilly, and Y. Bengio. "Deep learning needs a prefrontal cortex". *Work Bridging AI Cogn Sci* 107:603-616, 2020, p. 1.

180. K. Sakai. "Task set and prefrontal cortex". *Annu. Rev. Neurosci.* 31, 2008, pp. 219–245.

181. M. Sarafyazd and M. Jazayeri. "Hierarchical reasoning by neural circuits in the frontal cortex". *Science* 364:6441, 2019, eaav8911.

182. R. Schaeffer, M. Khona, and I. Fiete. "No free lunch from deep learning in neuroscience: A case study through models of the entorhinal-hippocampal circuit". *bioRxiv*, 2022, pp. 2022–08.

183. M. L. Schlichting and A. R. Preston. "Memory integration: neural mechanisms and implications for behavior". *Current opinion in behavioral sciences* 1, 2015, pp. 1–8.

184. W. Schneider and J. M. Chein. "Controlled & automatic processing: behavior, theory, and biological mechanisms". *Cognitive science* 27:3, 2003, pp. 525–559.

185. G. Schoenbaum, M. R. Roesch, T. A. Stalnaker, and Y. K. Takahashi. "A new perspective on the role of the orbitofrontal cortex in adaptive behaviour". *Nature Reviews Neuroscience* 10:12, 2009, pp. 885–892.

186. D. Shahnazian and C. B. Holroyd. "Distributed representations of action sequences in anterior cingulate cortex: A recurrent neural network approach". *Psychonomic Bulletin & Review* 25:1, 2018, pp. 302–321.

187. A. Shenhav, M. M. Botvinick, and J. D. Cohen. "The expected value of control: an integrative theory of anterior cingulate cortex function". *Neuron* 79:2, 2013, pp. 217–240.

188. A. Shenhav, J. D. Cohen, and M. M. Botvinick. "Dorsal anterior cingulate cortex and the value of control". *Nature neuroscience* 19:10, 2016, pp. 1286–1291.

189. A. Shenhav, M. A. Straccia, J. D. Cohen, and M. M. Botvinick. "Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value". *Nature neuroscience* 17:9, 2014, pp. 1249–1254.

190. M. Silvetti, R. Seurinck, and T. Verguts. "Value and prediction error in medial frontal cortex: integrating the single-unit and systems levels of analysis". *Frontiers in human neuroscience* 5, 2011, p. 75.

191. B. Skinner. *Principles of Behavior*. 1944.

192. C. Stevens. "Neural and cognitive bases of confirmation bias-induced interference in declarative memory performance". Theses. Université de Bordeaux, 2022. URL: https://theses.hal.science/tel-03641602.

193. A. Strock, X. Hinaut, and N. P. Rougier. "A robust model of gated working memory". *Neural Computation* 32:1, 2020, pp. 153–181.

194. D. Stuss, B. Levine, M. Alexander, J. Hong, C. Palumbo, L. Hamer, K. Murphy, and D. Izukawa. "Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: effects of lesion location and test structure on separable cognitive processes". *Neuropsychologia* 38:4, 2000, pp. 388–402.

195. R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

196. J. Tani and S. Nolfi. "Learning to perceive the world as articulated: an approach for hierarchical learning in sensory-motor systems". *Neural Networks* 12:7-8, 1999, pp. 1131–1141.

197. M. Todd, Y. Niv, and J. D. Cohen. "Learning to use working memory in partially observable environments through dopaminergic reinforcement". *Advances in neural information processing systems* 21, 2008.

198. N. Trouvain, L. Pedrelli, T. T. Dinh, and X. Hinaut. "Reservoirpy: an efficient and user-friendly library to design echo state networks". In: *International Conference on Artificial Neural Networks*. Springer. 2020, pp. 494–505.

199. J. K. Tsotsos and A. Rothenstein. "Computational models of visual attention". *Scholarpedia* 6:1, 2011, p. 6201.

200. M. T. Van Kesteren, D. J. Ruiter, G. Fernández, and R. N. Henson. "How schema and novelty augment memory formation". *Trends in neurosciences* 35:4, 2012, pp. 211–219.

201. T. Verguts. "Binding by random bursts: A computational model of cognitive control". *Journal of cognitive neuroscience* 29:6, 2017, pp. 1103–1118.

202. T. Verguts, E. Vassena, and M. Silvetti. "Adaptive effort investment in cognitive and physical tasks: a neurocomputational model". *Frontiers in Behavioral Neuroscience* 9, 2015, p. 57.

203. J. D. Wallis, K. C. Anderson, and E. K. Miller. "Single neurons in prefrontal cortex encode abstract rules". *Nature* 411:6840, 2001, pp. 953–956.

204. J. D. Wallis and E. K. Miller. "From Rule to Response: Neuronal Processes in the Premotor and Prefrontal Cortex". *Journal of Neurophysiology* 90:3, 2003, pp. 1790–1806. DOI: 10.1152/jn.00086.2003. URL: https://doi.org/10.1152%2Fjn.00086.2003.

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

206. J. X. Wang, Z. Kurth-Nelson, D. Tirumala, H. Soyer, J. Z. Leibo, R. Munos, C. Blundell, D. Kumaran, and M. Botvinick. "Learning to reinforcement learn". *arXiv preprint arXiv:1611.05763*, 2016.

207. C. J. C. H. Watkins. "Learning from delayed rewards", 1989.

208. A. Westbrook, B. Lamichhane, and T. Braver. "The subjective value of cognitive effort is encoded by a domain-general valuation network". *Journal of Neuroscience* 39:20, 2019, pp. 3934–3947.

209. I. M. White and S. P. Wise. "Rule-dependent neuronal activity in the prefrontal cortex". *Experimental brain research* 126, 1999, pp. 315–335.

210. R. C. Wilson, Y. K. Takahashi, G. Schoenbaum, and Y. Niv. "Orbitofrontal cortex as a cognitive map of task space". *Neuron* 81:2, 2014, pp. 267–279.

211. S. P. Wise. "Forward frontal fields: phylogeny and fundamental function". *Trends in neurosciences* 31:12, 2008, pp. 599–608.

212. J. R. Wolpaw and A. Kamesar. "Heksor: the central nervous system substrate of an adaptive behaviour". *The Journal of Physiology* 600:15, 2022, pp. 3423–3452.

213. J. N. Wood and J. Grafman. "Human prefrontal cortex: processing and representational perspectives". *Nature reviews neuroscience* 4:2, 2003, pp. 139–147.

214. N. Yeung, M. M. Botvinick, and J. D. Cohen. "The neural basis of error detection: conflict monitoring and the error-related negativity." *Psychological review* 111:4, 2004, p. 931.

215. E. A. Zilli and M. E. Hasselmo. "Modeling the role of working memory and episodic memory in behavioral tasks". *Hippocampus* 18:2, 2008, pp. 193–209.

216. D. Zipser, B. Kehoe, G. Littlewort, and J. Fuster. "A spiking network model of short-term active memory". *Journal of Neuroscience* 13:8, 1993, pp. 3406–3420.