Automatic and deliberate control of action: an embodied perspective of artificial and biological brains

Giovanni Maffei

TESI DOCTORAL UPF / ANY 2018

DIRECTOR DE LA TESI

Prof. Dr. Paul F.M.J. Verschure
Departament of Information and Communcation
Technologies

upf.
Universitat
Pompeu Fabra
Barcelona

To my family

Acknowledgements

I would like to sincerely thank all the persons that made this work possible. My gratitude goes to my supervisor Prof. Paul Verschure, for his guidance and vision. I am very thankful to my collaborators Ivan Herreros, Marti Sanchez-Fibla, Diogo Santos-Pata, Jordi Puigbo and all the colleagues from the SPECS Lab for the important contributions to my scientific and personal growth. A special mention goes to Lorenzo Rinaldi who is an example of scientific passion and brightness to me. Finally, my deepest gratitude and love is for Klaudia Grechuta, whose joy, strength and insight have made these years truly unique.

Abstract

Animals evolved to survive in dynamic environments by developing multiple behavioral strategies to adapt and to learn from their interaction with the world. Associative mechanisms and internal representations are at the core brain computation, however, to acquire a complete knowledge of their relevance for behavior it is necessary to take into consideration their embodied nature. In an interdisciplinary effort which integrates methods from computational modeling, robotics, and electrophysiology, this dissertation presents a series of studies that aim at advancing the understanding of the automatic and deliberate processes that regulate embodied control of action in the brain. Through the formulation of a biologically constrained control architecture engaged in a real-world foraging task, we lay the ground for modeling and analyzing complex goal-oriented behavior emerging from the interplay between the automatic cerebro-cerebellar system acquiring sensory-motor associations, and the deliberate fronto-hippocampal system providing goal-oriented navigation and planning. Following the behavioral analysis of the stimulus-response model of cerebellar learning, we later ask how could the cerebellum implement anticipatory control which is both adaptive and resistant to uncertainty. To answer this question, we explore the properties of the automatic control system and advance a novel hypothesis on the role of the cerebellum, by recasting its computation in the perceptual domain. Finally, we ask how the automatic and deliberate systems interact during unexpected situations that require a sudden change of plans. By analyzing the neural dynamics of the human frontal cortex in the control of deliberate action switch, we support the contribution of low-frequency oscillatory dynamics within this area to orchestrate behavior, based on internal representations of goals and rules. Altogether these results contribute to our understanding of how automatic and deliberate processes control action in the brain and advance novel insights that challenge or extend current theories. Despite the main aim to understand the brain, these insights could also be applied to the development of novel control systems for a new generation of robots.

Resum

Els animals van evolucionar per sobreviure en entorns dinàmics desenvolupant múltiples estratègies de comportament per adaptar-se i aprendre de la seva interacció amb el món. Els mecanismes associatius i les representacions internes estan en el nucli del càlcul del cervell, però, per adquirir un coneixement complet de la seva rellevància per a la conducta, cal tenir en compte la seva naturalesa incorporada. En un esforç interdisciplinari que integra mètodes de modelització computacional, robòtica i electrofisiologia, aquesta tesi presenta una sèrie d'estudis que pretenen avançar en la comprensió dels processos automàtics i deliberats que regulen el control de l'acció incorporat al cervell. Mitjançant la formulació d'una arquitectura de control biològicament restringida dedicada a una tasca de forjat en el món real, posem el terreny per modelar i analitzar una conducta orientada a objectius complexos que sorgeix de la interacció entre el sistema cerebrovascular cerebral automàtic que adquireix associacions sensorials motores i deliberat sistema fronto-hipocamp que proporciona una navegació i una planificació orientades a objectius. Després de l'anàlisi conductual del model d'estímul-resposta de l'aprenentatge cerebel lari, ens preguntem més endavant com es pot aplicar el cerebel a un control anticipat que és a la vegada adaptatiu i resistent a la incertesa. Per respondre a aquesta pregunta, explorem les propietats del sistema de control automàtic i avançem una nova hipòtesi sobre el paper del cerebel, tot reformant la seva computació en el domini perceptiu. Finalment, preguntem com interactuen els sistemes automàtic i deliberat durant situacions inesperades que requereixen un canvi sobtat de plans. Analitzant la dinàmica neural de l'escorça frontal humana en el control del canvi d'acció deliberada, recolzem l'aportació de dinàmiques oscil ladores de baixa freqüència en aquesta àrea per orquestrar el comportament, basant-se en representacions internes d'objectius i regles. Tot plegat, aquests resultats contribueixen a la nostra comprensió de com processos automàtics i deliberats controlen l'acció en el cervell i avançen noves idees que desafien o allarguen les teories actuals. Malgrat l'objectiu principal d'entendre el cervell, aquestes idees també es podrien aplicar al desenvolupament de nous sistemes de control per a una nova generació de

robots.

Publications

Included in the thesis

Peer-reviewed

Maffei, G., Santos-Pata, D., Marcos, E., Sanchez-Fibla, M., & Verschure, P. F. (2015). An embodied biologically constrained model of foraging: from classical and operant conditioning to adaptive real-world behavior in DAC-X. *Neural Networks, 72*, 88-108.

Maffei, G., Sanchez-Fibla, M., Herreros, I., & Verschure, P. F. (2014, July). The role of a cerebellum-driven perceptual prediction within a robotic postural task. In *International Conference on Simulation of Adaptive Behavior (pp. 76-87).* Springer, Cham.

Maffei, G., Herreros, I., Sanchez-Fibla, M., Friston, K. J., & Verschure, P. F. (2017, December). The perceptual shaping of anticipatory actions. In *Proc. R. Soc. B* (Vol. 284, No. 1869, p. 20171780). The Royal Society.

In preparation

Maffei, G., Puigbo, J. Y., Santos-Pata, D., Zucca, R., Principe, A., Roccamora, R., Conesa, G., and Verschure, P. (2018). Theta phase mediates deliberate action switch in human SMAs. (in preparation)

Other publications and abstracts

Peer-reviewed

Herreros, I., Maffei, G., Brandi, S., Sanchez-Fibla, M., & Verschure, P. F. (2013, November). Speed generalization capabilities of a cerebellar model on a rapid navigation task. In *Intelligent Robots and Systems (IROS), 2013*

IEEE/RSJ International Conference (pp. 363-368). IEEE.

Maffei, G., Sanchez-Fibla, M., Herreros, I., & Verschure, P. F. (2014, July). Acquisition of synergistic motor responses through cerebellar learning in a robotic postural task. In *Conference on Biomimetic and Biohybrid Systems* (pp. 202-212). Springer, Cham.

Ruck, M., Herreros, I., Maffei, G., Sanchez-Fibla, M., & Verschure, P. (2016, July). Learning to Balance While Reaching: A Cerebellar-Based Control Architecture for a Self-balancing Robot. In *Conference on Biomimetic and Biohybrid Systems* (pp. 214-226). Springer, Cham.

Maffei, G., Herreros, I., Sanchez-Fibla, M., & Verschure, P. F. (2013, July). Acquisition of anticipatory postural adjustment through cerebellar learning in a mobile robot. In *Conference on Biomimetic and Biohybrid Systems* (pp. 399-401). Springer, Berlin, Heidelberg.

Puigbo, J. Y., Maffei, G., Herreros, I., Ceresa, M., Ballester, M. G., & Verschure, P. F. M. J. (2018). Cholinergic Behavior State-Dependent Mechanisms of Neocortical Gain Control: a Neurocomputational Study. *Molecular neurobiology*, 55(1), 249-257.

Sanchez-Fibla, M., Maffei, G., & Verschure, P. F. (2017, September). Adaptively Learning Levels of Coordination from One’s, Other’s and Task Related Errors Through a Cerebellar Circuit: A Dual Cart-Pole Setup. In *International Conference on Artificial Neural Networks* (pp. 309-316). Springer, Cham.

Maffei, G., Puigbo, J. Y., & Verschure, P. F. (2017, July). Learning Modular Sequences in the Striatum. In *Conference on Biomimetic and Biohybrid Systems* (pp. 574-578). Springer, Cham.

Grechuta, K., Guga, J., Maffei, G., Ballester, B. R., & Verschure, P. F. (2017). Visuotactile integration modulates motor performance in a percep-

tual decision-making task. *Scientific reports*, 7(1), 3333.

Maffei, G., Herreros, I., Sanchez-Fibla, M., & Verschure, P. F. (2016, September). Plasticity in the Granular Layer Enhances Motor Learning in a Computational Model of the Cerebellum. In *International Conference on Artificial Neural Networks* (pp. 272-279). Springer, Cham.

Torgerud, W., Mussack, D., Lee, T., Maffei, G., Cotugno, G., & Schrater, P. Arousal both decreases conservativeness and increases attentiveness in a random dot motion decision making task. *Society for Neuroscience 2015*

Torgerud, W., Mussack, D., Lee, T., Maffei, G., Cotugno, G., & Schrater, P. Arousal and decision making: increased risk taking or improved attentiveness? *COSYNE 2015*

Contents

List of figures

Introduction

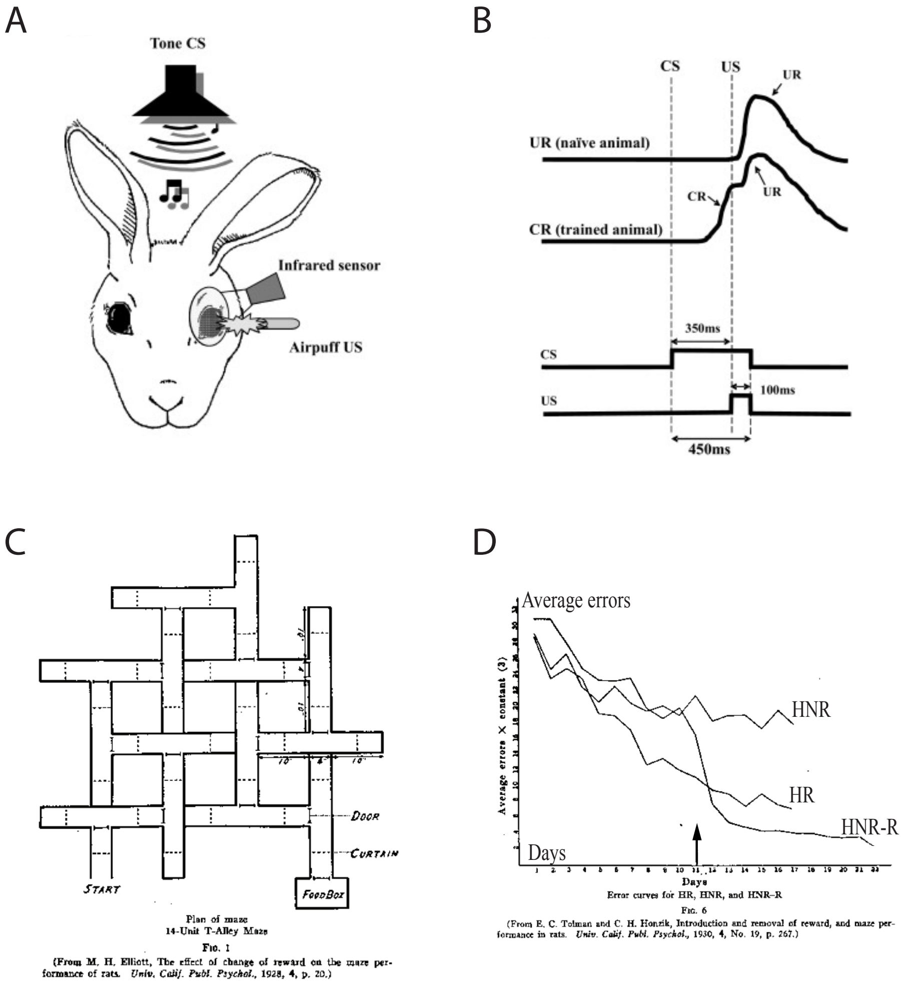
0.1 An embodied approach to the study of behavior

0.1.1 Historical context

The tradition of multiple approaches in the fields of psychology, cognitive science, and neuroscience have faced the challenge to describe and understand the dynamics and processes that underlie animal behavior. In the field of experimental psychology, early accounts of animal learning date back to the first half of the 20th century with the important contributions of Ivan Pavlov, Edward Thorndike, and Edward Tolman.

Pavlov introduced the notion of (Pavlovian) conditioning [Pavlov, 1941] to describe an experimental paradigm which taught animals to acquire new associations between stimuli and responses. This method to induce associative learning was based on a pairing between an Unconditioned Stimulus (US) and a novel one, termed Conditioning Stimulus (CS). A US represented a stimulus that instinctively, without training, provoked a reactive response of an animal (unconditioned response - UR). In Pavlov’s canonical experiment, the US consisted in the presentation of food which elicited salivation (UR) in the animal. The US was then paired with a CS such as a sound. Over time, the contiguity between the US and CS caused the CS to provoke the UR in the absence of the US. The newly associated response was termed the Conditioned Response (CR), to describe the new stimulus-response pair between the sound and the salivation. Another example of conditioned response can be found in the eyeblink conditioning paradigm, where an airpuff directed to the eyelid (US) provoking a blink (UR) is anticipated through the pairing with a tone (CS) (fig. 1-A,B) [Gormezano, 1972].

Edward Thorndike advanced an alternative associative paradigm by introducing the notion of consequence and demonstrated how animals could learn any type of stimulus-response associations through reward or punishment. This broader view of animal learning extended the conditioning paradigm from instinctual behaviors to genuinely novel ones and culminated in the formulation of the Law of Effect, the first psychological law of associative learning. Central to the Law of Effect was the idea that animals could learn any stimulus-response pair through training and repetition without forming an internal representation or a schema. Whereas the Law of Effect embraced and extended by the Behaviorism dominated the field of experimental psychology for over 50 years, Edward Tolman sought to demonstrate that animals were able to purposefully acquire knowledge about the environment, which they could subsequently use in a flexible manner, rather than learning stimuli-response associations only [Tolman, 1951]. Following the behaviorist experimental method, he observed that rats, who were trained to navigate a maze in search for food pellets, behaved in a more sophisticated manner than what the Law of Effect would predict. In particular, he suggested that animals were able to reuse knowledge of an environment previously acquired without reward. The concept of latent learning violated the essential principle of the Law of Effect demonstrating that learning can occur in the absence of reinforcement and inspired an alternative learning hypothesis grounded in representations (fig. 1-C,D). The notion of cognitive map described for the first time the fact that animals could make use of an internal representation of physical space acquired through experience and alluded to an internal schema upon which the animal could plan and deliberate [Tolman, 1938, Redish, 2016]. In sum, Tolman described learning as a consequence of a goal-oriented cognitive process that entails an internal representation of the environment in conjunction with motivational states. In a broader sense, this view on animal behavior stressed the existence of active cognitive processes that complemented the sensorimotor learning mechanisms predicated on the associationist view and enriched them with the notion of representation.



Plan of maze
14-Unit T-Alley Maze
FIG. 1
(From M. H. Elliott, The effect of change of reward on the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1928, 4, p. 20.)

Error curves for HR, HNR, and HNR–R
FIG. 6
(From E. C. Tolman and C. H. Honzik, Introduction and removal of reward, and maze performance in rats. *Univ. Calif. Publ. Psychol.*, 1930, 4, No. 19, p. 267.)

Figure 1: **Classical conditioning and Tolman's latent learning A.** Setup of the eyeblink conditioning paradigm. **B.** The pairing of CS and US progressively elicits a CR (from [Bracha et al., 1991]) **C.** The maze from Tolman's experiments **D.** Latent learning. Here rats that explored a

maze in absence of reward and that received reward starting from day 11 (arrow) (HNR-R) learned faster than those that received reward from day 1 (HR).(from [Tolman and Honzik, 1930])

0.1.2 Dual processor theories in psychology

The notion that multiple coexisting processes, based on associations and representations, may be essential for learning in animals is present in a wide range of theories formulated in the field of experimental psychology, macroscopically termed the Dual Processing System (DPS). At the core of DPS theories of mind lays the idea that the brain processes information in (at least) two ways including a fast and automatic one, and a slow and deliberate one. This may be found, for instance, in the definition of [Schneider and Shiffrin, 1977, Shiffrin and Schneider, 1977] who propose the coexistence of automatic and controlled processes in the context of visual attention. In their classical search experiment, automatic responses directed the attention towards the target stimulus, whereas controlled ones required conscious control of attention and memory. The first class of responses was considered to be a result of an acquired association of familiar stimuli, developed over a great number of repetitions. Controlled responses, on the other hand, were thought to involve an active information-processing mechanism which compares the presented stimuli with internal representations stored in memory during novel or infrequent situations [Schneider and Shiffrin, 1977, Shiffrin and Schneider, 1977, Schneider and Chein, 2003].

On a slightly different vein, [Sloman, 1996] defined two separate learning principles on the basis of human decision-making. Specifically, one associative, driven by experience and repetition, and one rule-based extracting knowledge through logical reasoning. The most popular example of DPS, however, can be found in the work of Kahneman, the winner of the 2013 Nobel prize. Based on the study of cognitive biases in human decision-making, Kahneman postulated the existence of the so-called System 1 and System 2. System 1 describes a set of processes, evolutionarily old, that underlie fast and automatic thinking and that take place at the unconscious level according to prior experience. Central to System 1 is the association of new information with existing patterns rather than creating new ones [Tversky and Kahneman, 1974]. System 2, in contrast, acts at a more controlled conscious and abstract level performing slower deliberate problem solving by taking decisions during, for example, novel situations [Kahneman, 2011].

In conclusion, DPS theories highlight two distinct mechanisms which underlie animal and human behavior. Consequently, they extend the early psychological framework of learning proposing that knowledge is acquired through processing of information by means of associations and internal representations.

0.1.3 An embodied perspective

In distinguishing between associations and representations both early and more recent accounts of animal learning stressed the idea that central to these processes is the ability of the brain to acquire, process and output information. This is most evident in the theoretical frameworks attempting to capture the computational aspects at the basis of classical and operant conditioning [Sutton and Barto, 1998, Rescorla et al., 1972].

For example, the Rescorla-Wagner rule (1972), developed by Robert Rescorla and Allan Wagner to formalize the principles behind the conditioning phenomena, interpreted associative learning in terms of an acquired internal organization of memories, formed of relationships between events across a variety of modalities. [Rescorla et al., 1972]. Here however, an association is not just an encoding of a co-occurrence of two stimuli (CS-US) but rather an expectation that establishes a causal link between related events in the world [Rescorla, 1988, Medin et al., 1996, Spear, 2014]. This aspect is also captured by early attempts to implement conditioning dynamics into artificial neural networks [Donegan et al., 1989] and later extensions to probabilistic normative frameworks [Chater et al., 2006]

Moreover, the emphasis on memory and internal representation is present in a wide range of cognitive theories of decision-making and deliberation that distinguish different learning modes based on the different information processing capabilities of distinct areas of the brain [van der Meer et al., 2012, Oppenheimer and Kelso, 2015]. For example, theoretical interpretations of the original concept of cognitive map describe the formation of internal representations as an unsupervised process which captures the statistics of redundant flows of sensory inputs [Barlow, 1989]. An internal schema that can be further used to perform model-based planning through the representations of goals, actions, and expected outcomes [Daw, 2012].

Finally, a similar focus on the information processing aspects of learning is at the core of the formulation of dual processing theories of cognition. In particular, a paradigmatic example can be found in the implementations of the theory of Schneider and Shiffrin into a computational model which aims at specifying the structures underlying cognitive performance in terms of information encoding and retrieval [Schneider and Chein, 2003]. Here, the information available in the environment is processed by an automatic associative system and a controlled one composed of intentional and mnemonic processing modules. Both systems alter the input in a systematic, purposeful way in order to generate the desired output.

In a broader sense, the notion of information processing is central to the field of cognitive science [Fodor, 1985] and its relation to the computational theory of mind advanced by connectionist approaches as an algorithmic description of cognitive functions [Jackendoff, 1987]. A view which is further promoted by current interdisciplinary approaches attempting to bridge the fields of neuroscience and artificial intelligence based on a close analogy between biological and artificial (deep) neural networks [Hassabis et al., 2017, Dehaene et al., 2014] or that describe mental operations in terms of unified algorithmic principles that extract probabilistic rules that govern the environment minimizing global objective functions [Tenenbaum et al., 2011, Friston, 2010].

Throughout this dissertation, we will argue, however, that focusing exclusively on the information processing aspects of learning might constrain the understanding of the mechanisms that underlie biological behavior. This is because the brain is not an abstract information processor but rather an embodied system whose primary goal is to control the body in its continuous interactions with the environment [Clark and Squire, 1998, Varela et al., 2017]. This becomes apparent from the perspective of the evolution of the Nervous System (NS). Simple NSs have evolved into more complex and structured ones in order to provide reactive and adaptive capabilities to organisms equipped with increasingly articulated bodies which had to act within ever-changing, competitive environments [Jerison, 2012, Kaas, 2013]. For example, in mammals the evolution of the nervous system accompanied the transition to novel ecological niches where to forage [Simpson, 1955]. This required the development and expansion of new peripheral apparatus and associated brain structures to handle more complex coordination, but also new homeostatic mechanisms to regulate, for example, body temperature. Corollary to this is the basic proposition that the amount and kind of nervous tissue in the brain is related to the amount and kind of sensorimotor activity that must be controlled [Jerison, 1971], with increasingly complex brains representing sensorimotor interactions with the environment in a more sophisticated and abstract way [Fuster, 2015, Stout, 2010, Jerison, 2012].

This has implications for the dual processing system theories because in the context of an embodied system the brain should act not just as an information processing system but also as a control system. Under this new perspective the learning abilities of the brain acquire a boarder meaning: learning and adaptation have the ultimate evolutionary goal of regulating the sensorimotor interactions of an individual with the environment for the purposes of achieving goals that satisfy an internal need that come from the body (i.e. nutrition) [Damasio and Sutherland, 1994]. Further, the content of the information processed by the brain is grounded in the sensorimotor space and it is processed with the goal to act in the world. Therefore, information processing mechanisms, in their diversity, should be seen as a whole in the context of controlling a physical body. In this sense, more complex and abstract representations should stem from an incremental development from simpler mechanisms through a bottom-up evolutionary process and they are not based on centralized modules but rather emerge from the integrated functions of distinct sub-modules within the distributed (nervous) systems. Failing to integrate this perspective might incur in the misconception that cognition consists simply of building maximally accurate representations of input information, while learning is a mean to achieve a more immediate goal of guiding behavior in response to the system’s changing surroundings (see [Miłkowski, 2013], p. 4). For this reason to advance our understanding of automatic and deliberate processes within this dissertation we focus on the study of action. In particular we will attempt to complement the information processing approach with the notion of embodied control. By looking at the brain as a control system that acts within the environment, we will propose a different view on automatic and deliberate processes. Not only we will demonstrate their capabilities to acquire associations and representations but also their function in the purposeful control of mechanical and biological bodies. In doing this, we will provide a theoretical framework, presented in part 1, that maps the functional description of automatic and deliberate processes to the neural substrate. In doing this we will advance the understanding of the neurobiological origin of these processes and generate a number of testable hypotheses about the individual subsystems, which will be developed in the following chapters.

0.1.4 The brain as a control system

The approach to the study of the brain as an embodied control system has a long-standing tradition grounded in the cybernetics of Wiener, Ashby, and Walter [Ashby, 1949, Wiener, 1948, Walter, 1963]. The field of Cybernetics was born in the late 1940s, the same historical context in which connectionism emerged, as an interdisciplinary field whose goal was to study the design principles as well as mechanisms underlying control and communication of biological and mechanical systems in their interaction with the environment [Wiener, 1961].

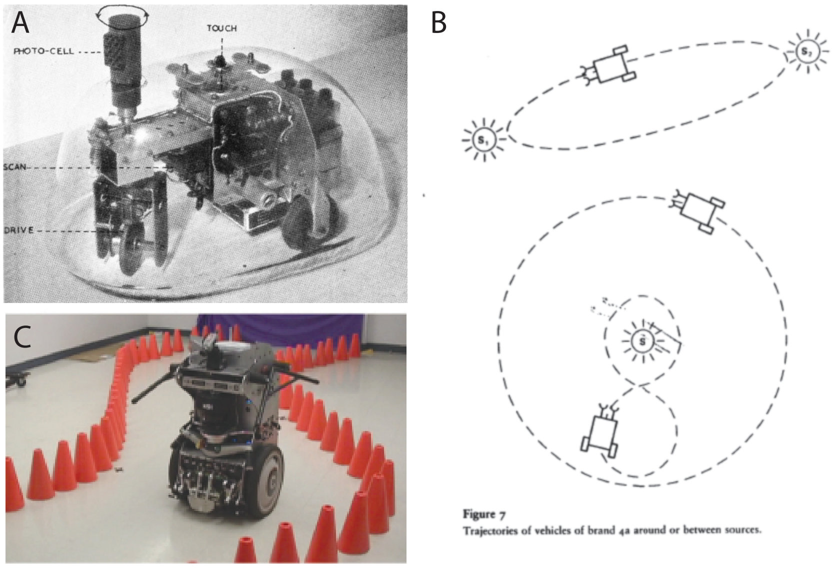


Figure 2: **Cybernetic approach to the study of behavior A.** the tortoise from Gray Walters **B.** An illustration of complex behavior in the Vehicles (from [Braitenberg, 1986]) **C.** A brain based device navigating a track (from [Edelman, 2007])

The concept of information processing was central to both connectionism and cybernetics. The latter, however, stressed its application on control rather than computation. In particular, Cybernetics introduced a conceptual framework to understand biological systems grounded in the notion of feedback. Here, the system creates a closed loop with the environment based on perception and action cycles under the general hypothesis that the main function of the brain is to control, in a stable way, the behavior of the animal (i.e. homeostatic functions [Ashby, 1949]). Methodologically, Cybernetics promoted interdisciplinary research which addressed integrative questions about the mechanisms of control in animals and machines through the construction of robotic models. A notable example can be found in the tortoises of Gray Walters, small autonomous robots built with the goal to demonstrate that complex behavior (such as phototaxis) could emerge from a relatively simple mapping between perceptual (sensors) and motor (actuators) systems (fig. 2-A).

The field of Cybernetics has been historically surpassed by connectionist approaches possibly due to the limitations of building complex machines instead of writing computer programs. The principles promoted by the cybernetic approach, however, were further developed in neuroscience and psychology. Such principles, for example, can be found in the methods of Valentino Braitenberg, a psychiatrist, and neurophysiologist who coupled the study of the biological brain with theoretical insights borrowed from computer science and engineering [Braitenberg, 1990, van Hemmen et al., 2014]. In his book entitled Vehicles, Braitenberg proposes a thought experiment in which simple agents display simple intelligent behaviors based on pure feedback and without the need for complex internal representation [Braitenberg, 1986] (fig. 2-B). The emphasis on intelligence without representation is perhaps the central tenet of the subsumption architecture of Rodney Brooks who, as an answer to the symbolic doctrine, introduced an approach for the design of Artificial Intelligence (AI) grounded in embodied systems [Brooks, 1991].

Finally, the Nobel prize, Gerald Edelman, advanced the so-called ’Darwin series’ of real-world Brain-Based Devices (BBD) arguing that such synthetic method is the key step to improve our understanding of how behavior is generated by the system-level interaction of multiple artificial neuro-mimetic controllers [Edelman, 2007]. The assumption on top of which the BBD where built was that the brain is an embodied system that cannot be fully understood outside of the context of the interaction between brain, body, and environment. This line of research led to a construction of a number of different artifacts which implemented neural models of different brain structures. Indeed, those artifacts were capable of adaptive behaviors in the real world including goal-oriented navigation as well as conditioned responses to stimuli, in a situated way, without global coordinate systems or priors but just through sensing and actuating (fig. 2-C).

0.1.5 Mapping brain functions to layers of control

In this dissertation we will follow the embodied approach to the study of behavior and look at brain mechanisms underlying automatic and deliberate processes in the control of action. In order to ground our definition of such processes, we introduce the notion of brain architecture. This notion allows the double purpose of (1) defining the design principles that an embodied system should follow in order to purposefully control behavior and (2) mapping those principles to the biological neural substrate. In particular, we frame our definition into the Distributed Adaptive Control framework (DAC) [Verschure et al., 2003b, Verschure, 2012]. DAC extends the tradition of Edelman’s BBD and describes the functional and anatomical organization of the mammalian brain as divided into four complementary layers of control:

- *somatic layer*: it describes the fundamental interface between an embodied agent and its environment. From a control perspective it defines the physical aspects of the controlled system (i.e. the plant) including its sensors and actuators as well as internal states. From a biological perspective the somatic layer describes the body and comprises the senses, the skeleton-muscle system as well as, for example, the circulatory and digestive systems.

- *reactive layer*: it describes reactive behaviors, such as reflexes (US-UR), and homeostatic mechanisms. From a control perspective it defines hardwired feedback control mechanisms that map the sensors to the actuators and produce stereotyped responses to sensory inputs, for example to minimize a perceived error. From a biological perspective the reactive layer describes the reflex arc and homeostasis and it has been mapped to the functions of the brainstem and the hypothalamus in the mammalian brain. Here the hypothalamus is involved in, for example, thermoregulation whereas areas such as the pons, in the brainstem, provide the mapping between sensory and motor fibers underlying, for example, eyblink reflexes.

- *adaptive layer*: it describes adaptive behaviors based on associative mechanism such as stimulus-stimulus and stimulus-response associations (i.e. CS-CR). From a control perspective this defines adaptive feedforward control, steer the actuators on the basis of previous knowledge of the controlled system, for example to minimize an error before it is perceived. From a biological perspective the adaptive layer defines the associative learning system and it has been ascribed to the functions of the cerebellum [Herreros and Verschure, 2013a, Herreros et al., 2013a, Maffei et al., 2013, Verschure, 2012, HofstoÈtter et al., 2002]. Here, sensory and motor afferents from distinct areas of the brain are mapped according to detected contingencies through plasticity mechanism and allow for anticipatory behaviors such as avoiding a noxious stimulus (see section 2). Other structures within this layer are, for example, the striatum in its sensorimotor functions associated to outcomes.

- *contextual layer*: it describes goal-oriented behaviors based on deliberation, that makes use of internal representations of goals, space and value (i.e. latent learning). From a control perspective it defines control of action based on planning. Planning is defined as the ability to act flexibly, by switching between the systems objectives according to contextual changes and internal states and to accomplish those objectives upon consultation of memory and expected outcomes. From a biological perspective, the contextual layer represents the deliberate learning system and it has been mapped to the functions of the frontal cortex, where abstract representations of actions are flexibly orchestrated in relationship to goals [Verschure et al., 2014a, Duff et al., 2011] (see section 3). This system also involves distinct memory areas, such as the hippocampus and the temporal lobe, as well areas associated with value, such as the ventral striatum and the ventral tagmental area.

This definition is, by no means, incomplete and it is sacrificing the richness and complexity of the nervous system in favor of a schematic representation of the brain in distinct layers of control. Nevertheless it captures the basic principles that underlie biological behavior and it allows to develop concrete hypotheses on the interactions between distinct layers and biological subsystem in the control of an embodied system.

Therefore, within this dissertation, we ground our definition of automatic system within the reactive and adaptive layers of control and, specifically, into the functions of the cerebellum. In turn, we ground our definition of deliberate system within the contextual layer and in particular into the functions of the frontal cortex. In section 2 and 3 of this chapter we will cover in more detail the basic anatomical and neurophysiological features of these two areas and theoretical accounts that describe their involvement in controlling behavior.

0.2 Automatic control in the Cerebellum

0.2.1 Anatomical organization of the cerebellum

The cerebellum is located in the posterior part of the brain overlying the brainstem at the level of the pons (fig. 3-A). It is one of the oldest structures of the nervous system of the mammalian brain. The earliest anatomical mapping of the cerebellum dates back to the Spanish physiologist Ramon y Cajal who, already in 1911, identified its main cell types through histological staining [Ramón y Cajal, 1911]. A complete characterization of its anatomical and physiological properties emerged in 1967 with the work of Eccles, Ito, and Szentagothai, who established its organization in a defining formulation still accepted nowadays [Eccles et al., 1967]. The mammalian cerebellum receives inputs from a widespread of cortical and sub-cortical regions through the pontine nuclei and its efferent connections project to brainstem motor centers and especially back to the cortex in an organization often referred to as cortico-cerebellar loops [Middleton and Strick, 1998]. In mammals, the cerebellum expanded at the same rate of the cortex suggesting a tight functional coupling between these two structures [Jerison and Barlow, 1985].

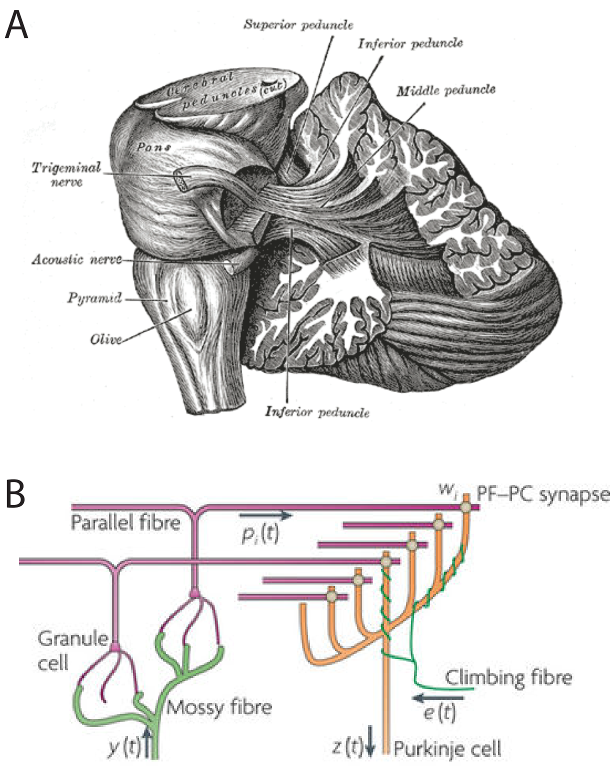


Figure 3: **Cerebellum and cerebellar microcircuit A** Section of the human cerebellum and pons (from [[Warwick et al., 1973]]) **B** Schematic representation of the cerebellar microcircuit (from [[Dean et al., 2010a]])

The cerebellum is cytoarchitectonically divided into the cerebellar cortex and the cerebellar nuclei. The cerebellar cortex can be further decomposed into granular and molecular layers. The former is located in the innermost part of the cerebellar cortex and it is mostly composed of the granule cells that receive inputs from the pons via mossy fibers and relay this signal to parallel fibers in the molecular layer [Eccles et al., 1967, Ito, 2006, Apps and Garwicz, 2005]. The main cell type within the molecular layer is the GABAergic Purkinje cell. Purkinje cells are among the largest cells known in the mammalian brain and they are characterized by a large arborization of the dendritic tree making hundreds of thousands of synapses with the parallel fibers.

Purkinje cells axons represent the only output of the cerebellar cortex onto the deep cerebellar nuclei. The cerebellar nuclei are composed of constantly active excitatory and inhibitory cells that, through disinhibition from the Purkinje cells, project the output signal of the cerebellum to their target structures [Uusisaari and de Schutter, 2011]. Another structure that anatomically relates to the cerebellum is the inferior olivary nucleus in the medulla [Brown et al., 1977]. It receives afferents from sensory-motor subcortical and cortical areas and projects to the dendritic tree of the Purkinje cells through the climbing fibers [Eccles et al., 1967, Ito et al., 1982]. The cerebellum presents a strikingly repetitive structure where the Purkinje cells and associated cell types are organized in a set of parallel circuits often referred to as ”cerebellar microcircuits” (fig. 3-B) [Eccles et al., 1967, Braitenberg and Atwood, 1958]. Although a number of other cell types have been characterized within the cerebellar microcircuit, this simplified description provides an anatomical foundation for the validation of an algorithmic hypothesis of its functions [Apps and Garwicz, 2005, Paul Dean and John Porrill, 2010].

0.2.2 What does the cerebellum compute?

One of the main models describing the computation implemented in the cerebellum is based on the Albus-Marr hypothesis [Marr, 1969, Albus, 1971]. Starting from a functional assumption that the cerebellum was related to motor control on the basis of anatomical, clinical and lesion evidence available at the time, Albus and Marr hypothesized that the cerebellar microcircuit, and in particular the Purkinje cell, could learn associations between sensory states (i.e. context) and motor commands to be issued in that context. This interpretation was supported by the anatomical input-output connectivity of the cerebellar microcircuit, which would receive contextual (sensory) information through the pontine pathways and the motor command (or pattern) to be learned through the input from the climbing fibers in the form of a supervising signal [Eccles et al., 1967].

Following this initial (discrete) formulation, further refinement of the theory proposed the analogy of the cerebellum as an adaptive filter [Fujita, 1982]. Adaptive filters are a class of linear filters where the input-output transfer function parameters can be adjusted via an optimization method [Widrow et al., 1967]. Computationally, an adaptive filter requires a two-steps analysis-synthesis procedure. During the analysis step, an input is expanded into a set of basis functions with different temporal profiles, whereas during the synthesis step an output is computed via a weighted sum of those bases. The weights used to mix the bases into a unitary signal are adjusted with the goal to minimize the difference between the output of the filter and a target function (i.e. least square error).

According to this analogy [Fujita, 1982,Dean et al., 2010a], the analysis step would occur in the granular layer where the interaction between the granule cells and the Golgi cells could provide the necessary transformation of the input signal into temporal bases, further conveyed to the Purkinje cell dendrite through the parallel fibers. The Purkinje cell would, therefore, compute the synthesis step of the algorithm by integrating parallel fiber inputs into a unique response. Plasticity at the parallel fibers - Purkinje cell synapsis serves as a weight adjustment regulated by Long-Term Depression (LTD) (however see [Johansson et al., 2014]) driven by an error signal computed at the level of the inferior olive [Ito et al., 1982, Hesslow, 1994].

Overall, the hypothesis that the role of the cerebellum (i.e. internal computation) is to learn, reproduce and perhaps anticipate (i.e. eligibility trace) an arbitrary function provided the right set of inputs and outputs, has been supported by anatomical, behavioral and neurophysiological evidence (see [Dean and Porrill, 2011] for review). Nevertheless, the original model underwent a progressive refinement and revisions that enriched it with new plasticity sites on the basis of new behavioral and anatomical observations (see [Clopath et al., 2014] for an example). What surprisingly still remains elusive about the cerebellum, however, is its functional role with respect to the rest of the brain [Ito, 2006, Ito, 2008, Ramnani, 2006b].

0.2.3 Anticipatory behaviors in the cerebellum

The highly repetitive structure of the cerebellum, organized as microcircuits, suggests an algorithmic homogeneity of its internal computation [Eccles et al.,] (although see [Cerminara et al., 2015]). What contributes to determine the functions of the cerebellum, however, is the input-output connectivity with the rest of the brain, as suggested by the wide range of behaviors it is involved in.

The cerebellum has been implicated in a range of motor and non-motor functions that have determined its classical parcellation into vestibular, spinal and cerebral [Ito, 2008, Ramnani, 2006b]. Perhaps the most established behavior that involves cerebellar learning is the adaptive eyeblink conditioning reflex [Gormezano, 1972] (fig. 1-A,B). A similar anticipatory dynamic is found in the smooth pursuit behavior where the cerebellum drives a predictive response that allows the eyes to closely follow a moving target [Shidara et al., 1993], or in the vestibular-ocular reflex where the cerebellum elicits an anticipatory eye movement that compensates for the head rotation [Fukuda et al., 1972, Miles and Lisberger, 1981]. The involvement of the cerebellum extends to the control of limbs as in the force-field adaptation task where it is critical to anticipate and minimize the effects of a disturbance during a reaching movement [Pasalar et al., 2006, Shadmehr and Mussa-Ivaldi, 1994, Thoroughman and Shadmehr, 2000].

Finally, several empirical observations involve the cerebellum in the control of corrective postural adjustments [Massion, 1994, Timmann and Horak, 2001a, Lang and Bastian, 1999]. This behavior will be of particular interest within this dissertation and we briefly review here some essential concepts. Postural adjustments are defined as predictive responses issued by the brain to counteract the effect of an incoming disturbance affecting the state of equilibrium of the body before an error is perceived. This is experimentally observed when a quietly standing subject is hit by an obstacle [Santos et al., 2010b, Santos et al., 2010a] (fig. 4-A) or the platform on top of which he stands is moved forward [Timmann and Horak, 2001a].

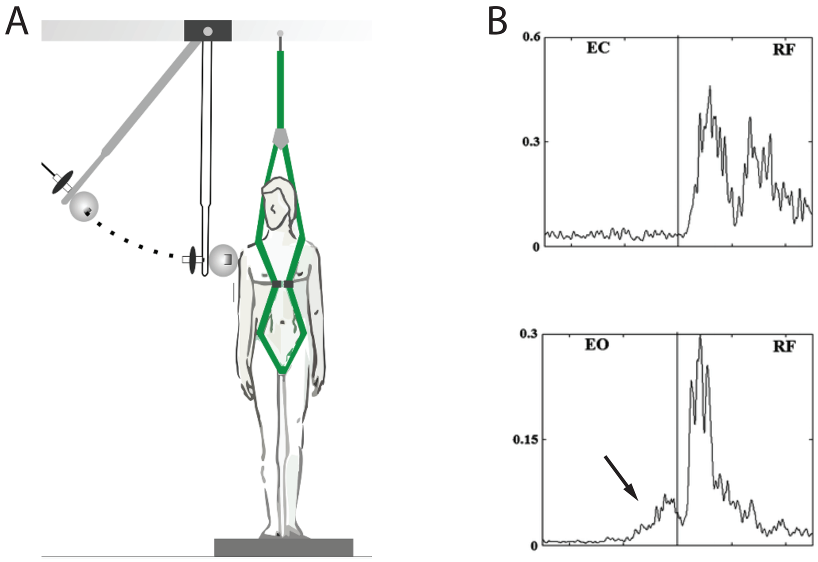


Figure 4: **Anticipatory postural adjustments A** Setup to study postural adjustments against external disturbances (from [Santos et al., 2010b]).**B** Example of electromyographic response recorded from the rectus femoris muscle of a subject resisting a disturbance with eyes closed (top) or eyes open (bottom). Only when the incoming disturbance can be predicted APAs are detected (from [Mohapatra et al., 2012]).

In both cases, the subject perceives a loss of equilibrium (i.e. error) to be compensated for and makes use of predictive cues (i.e. the proximity of the obstacle or a tone preceding the platform motion) to act anticipatorily (fig. 4-B). The total postural corrective response has been characterized as the blend of at least three components, possibly originated as the result of three distinct processes. In particular, the anticipatory component of the response precedes the moment of the impact, the early (or fast) compensatory component follows the moment of the impact but rises earlier than the time of a postural displacement is being perceived by the vestibular system (¡200ms). Finally, late (or slow) compensatory responses follow the perception of the vestibular error [Shiratori and Latash, 2001, Latash, 2008]. In conclusion, the cerebellum has been involved in a number of behaviors that require the anticipation of motor responses to counteract predictable errors and allow for fast and responsive actions within a number of domains.

0.2.4 Theories of cerebellar control

From a system-level modeling perspective, theoretical accounts supporting this plurality of behaviors can be traditionally divided into two classes of models. In particular, the ’inverse’ and the ’forward’ models [Wolpert et al., 1998b]. Both classes share the same building blocks capitalizing on the notion that the cerebellum provides a feedforward anticipatory signal. They differ, however, in the quality of ’what’ they predict. Both require the presence of a plant that is the system to be controlled

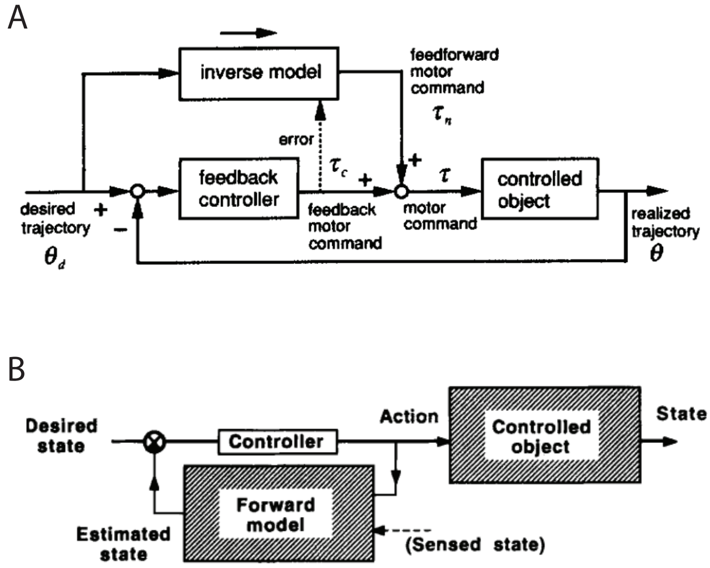


Figure 5: **Theoretical functions of the cerebellum A.**The inverse model hypothesis (from [Gomi and Kawato, 1992]) **B.**The forward model hypothesis (from [Miall et al., 1993])

(i.e. the eye or the limb) and a (feedback) controller that is in charge to achieve or maintain a desired state of the plant by mapping an error (i.e. difference between current and desired state) into a corrective control signal (i.e. the reflex). However, since the sensory feedback is subject to delays and noise and the dynamics of the plant may introduce further latencies (i.e. viscoelastic plants such as muscles), the feedback controller alone cannot fully correct for the error. For this reason, a feedforward controller that takes the function of the cerebellum is necessary to compensate for these behaviorally relevant control latencies.

Inverse models, such as Feedback Error Learning (FEL) [Kawato et al., 1987, Gomi and Kawato, 1992] (fig. 5-A), hypothesize that the cerebellum learns to produce a necessary motor command to achieve a desired state of the plant, covering a similar function of the feedback controller but in a predictive manner. In particular, FEL learns to associate a contextual or a predictive signal with a response that mimics the output of the feedback controller advanced in time with the function to minimize an error before it can be perceived through sensory feedback. This scheme has been used to describe behaviors such as the eyeblink conditioning [Christian and Thompson,], postural adjustments and VOR [Gomi and Kawato, 1992].

Forward models, such as in the Smith predictor analogy [Miall, 1998] (fig. 5-B), hypothesize that the cerebellum learns to predict the sensory consequences of motor commands to compensate for the latencies of sensory feedback. This is achieved on the basis of an internal model which describes the dynamics of the controlled object that converts an efference copy of a given motor command into a feedforward prediction about the incoming sensory feedback. This scheme is consistent with a number of (voluntary) motor behaviors including the smooth pursuit [Kettner et al., 1997] and force-field adaptation [Tseng et al., 2007, Kawato, 1999] (but see [Shadmehr et al., 2016]).

0.2.5 Open questions

The cerebellum is one of the best-characterized structures of the brain. Its main anatomical and electrophysiological features have been established and a number of behaviors dependent on them have been extensively studied and interpreted using the formalism of inverse and forward models. Nevertheless, a number of questions regarding the cerebellum remain open and need to be addressed by current research agendas, specifically about its anatomical details, physiology and its role with respect to the rest of the brain.

For the scope of this dissertation, we focus on two issues that regard the current theories of cerebellar functions. Theoretical control models of the cerebellum, as the examples mentioned above, have focused mostly on the analysis of one individual microcircuit or collapsed the functions of one cerebellar microzone [Ito, 1987] within a unitary algorithm. However, the cerebellum is composed by approximately 15 millions microcircuits with very different connectivity to the rest of the brain. In some cases this multiplicity has been modeled by parallel architectures that integrated the output of multiple microcircuits into a unitary response, capitalizing on the parallel nature of the cerebellar cytoarchitecture [Wolpert et al., 1998a, Wolpert and Kawato, 1998, Brandi et al., 2013]. Some other cases have emphasized the hierarchical scheme suggested by the multiple layers of the motor control hierarchy to which the cerebellum contributes [Kawato et al., 1987]. However, it remains an open questions how multiple microcircuits integrate their anticipatory functions, especially during the control of behaviors that require the blend of multiple adaptive responses, as in the case of postural adjustments. Indeed these responses could be the result of a parallel cerebellar scheme. Alternatively, they could stem from a sequential architecture. In these regards multiple possibilities should be explored and their implications for the control of behavior carefully studied.

A second aspect regards the function of the cerebellum with respect to the rest of the brain. The study of the cerebellum has a long-standing tradition within the realm of motor control and its role in motor behavior is widely accepted. However, a growing body of empirical evidence challenges the purely motor role traditionally ascribed to the cerebellum and broadens the functional involvement of this structure in different domains [Ramnani, 2006a, Strick et al., 2009a, Caligiore et al., 2017]. In particular, the cerebellum seems to play an important role in situations that involve the prediction of forthcoming sensory events even in cases where action is not required [Roth et al., 2013a, Deluca et al., 2014, Therrien and Bastian, 2015]. Anatomical evidence further supports the engagement of the cerebellum in purely sensory and cognitive domains, drawing connectivity patterns that reciprocally link it to different neocortical regions involved in sensory processing, such as somatosensory and parietal cortices [Strick et al., 2009b, Schmahmann, 1996]. Crucially this set of evidence is at odds with either the classical forward or inverse model formulations and rises the question of whether the multiplicity of behaviors the cerebellum is involved in are controlled by distinct learning schemes or perhaps there could be a unified view. If so, how could pure sensory predictions play a role in motor adaptation?

0.3 Deliberate control in the Frontal Cortex

0.3.1 Frontal control of goal-oriented behavior

The frontal lobes, the largest of the four major lobes of the cerebral cortex, are located in the frontal area of the mammalian brain, rostral to the central sulcus, an anatomical landmark that macroscopically separates the sensory and the motor regions of the neocortex. The frontal lobes are anatomically divided into the primary Motor Cortex (M1) and, more anteriorly, the premotor region, further segmented into the lateral premotor cortex and, centrally, the Supplementary Motor Complex (SMC). Anterior to the premotor region sits the prefrontal cortex, an area especially developed in primates and humans [Fuster, 2015]. The regions of the frontal lobes are interconnected by a feedforward communication pathway that allows information to flow from prefrontal to primary motor cortices and a feedback pathway that backprojects form motor areas to prefrontal ones [Fuster, 2015] (fig. 6).

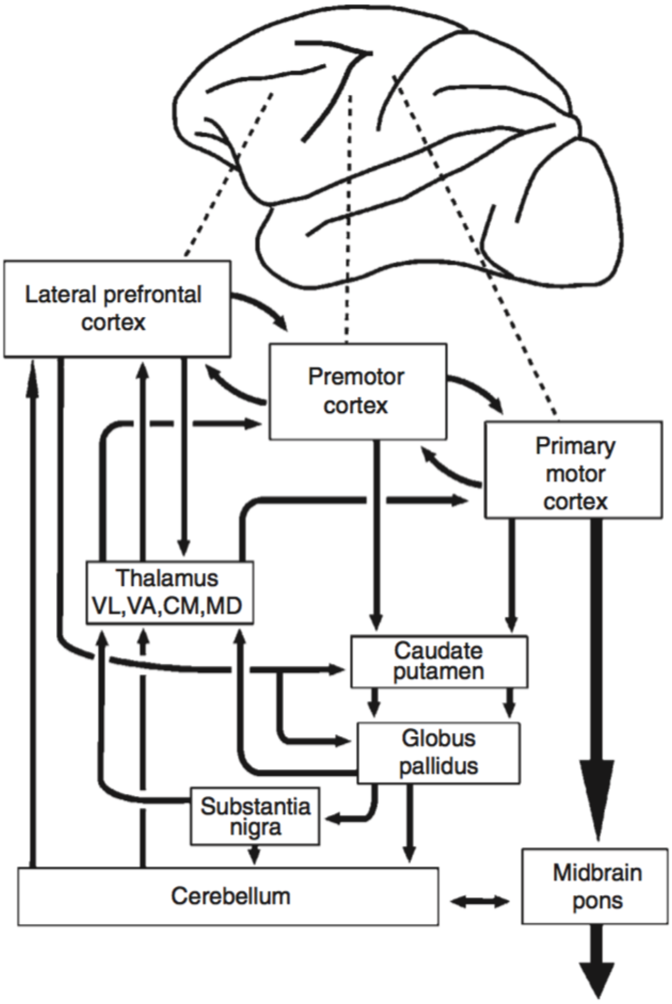


Figure 6: **Frontal networks involved in the control of action** (from [Fuster, 2015])

The frontal cortex, however, is also extensively connected to subcortical areas involved in sensory, memory and motor functions such as the basal ganglia, the hippocampus, the cerebellum and low-level motor nuclei in the brainstem, to name a few. Following the description proposed by Fuster [Fuster, 2015], the entirety of the cortex of the frontal lobe is devoted to the representation and production of action at all levels of biological complexity in a hierarchical organization that sees the most elementary actions operating at the low levels (i.e. M1) and the most complex and abstract actions sitting at the higher level of the hierarchy (i.e. lateral prefrontal cortex - LPFC). This definition entails, therefore, a progressive gradient from abstract goals (i.e. high-level description of behavioral patterns) to their motor implementations in the muscular-skeletal system. The involvement of the frontal lobes in cognitive control, that is the ability to coordinate actions in relation to internal goals, and their underlying hierarchical organization has been a subject of a number of studies [Koechlin, 2016, Ridderinkhof et al., 2004, Miller and Cohen, 2001, Nachev et al., 2008]. For example, using a cognitive task involving the contribution of sensory, contextual and episodic elements, [Koechlin et al., 2003] differentiated the contributions to the increased bold signal of human premotor and caudal and rostral lateral prefrontal cortex. The emerging model is a cascade of processes that mediate behavior in a hierarchical fashion where sensory evidence is processed in premotor areas whereas contextual and episodic control would be a distinctive feature of prefrontal areas. Similarly, [Voytek et al., 2015] suggested a progressive involvement of prefrontal cortex during a task that required the processing of increasingly abstract rules. Several other reports have supported the involvement of frontal areas in hierarchically organized functions such as decision-making, planning, and action-outcome monitoring, aspects central to the notion of cognitive control [Ridderinkhof et al., 2004].

0.3.2 Change of plans: the involvement of SMA

One of the key requisites for a ’cognitive controller’ in charge of executing goal-oriented actions is the ability to switch between alternative objectives following task demands and implement them at the level of the motor system. For example, by suppressing ongoing automatic behavior and overwriting it with a controlled ’deliberate’ action [Koechlin, 2016, Hikosaka and Isoda, 2010a, Nachev et al., 2008]. This is a key behavioral function of a goal-oriented system with great implications for survival as, for instance, when driving a car and unexpectedly having to steer because of an obstacle. From an experimental perspective, it also depicts an extremely valuable case to study to understand the interaction between automatic and deliberate (or controlled) processes through the creation of a conflict, in which the latter has to exert control over the former. The switch from automatic to deliberate processes is behaviorally characterized phenomenon that can be observed during ’change of plans’ experimental paradigms [Isoda and Hikosaka, 2007, Rushworth and Hadland, 2002]. This class of paradigms requires a subject to execute a repetitive action or a set of actions in response to a stimulus (i.e. context) which induces a progressive shift from reactive responses, where action is guided by the cue, to anticipatory, or feedforward responses, where action is guided by predictive processes that anticipate the presence or location of the stimulus. This progressive automation of the response is captured by the gradual decrease in stimulus-response time. On a small subset of trials, however, an unpredictable change of contingencies requires the subject to switch from an over-trained goal to a novel (or less frequent) one. Switch trials are typically accompanied by a higher number of mistakes (i.e. not being able to switch) and, importantly, longer response times, which suggests the intervention of a slower and perhaps more complex system in the control of action [Isoda and Hikosaka, 2007].

0.3.3 Supplementary motor areas: a neural controller for deliberate action switch?

A candidate area for the implementation of goal-oriented plans into the motor system is the Supplementary Motor Complex (SMC). Its anatomical location indeed sits in between the prefrontal system and the primary motor cortex, suggesting a possible functional interface between high-level deliberation and actuation [Nachev et al., 2008]. The involvement of the SMC in motor control was originally observed during stimulations studies. They implicated SMC in the control of complex behavior such as bimanual coordination or production of motor sequences and described its functions as auxiliary to the primary motor system [Penfield and Jasper, 1954]. These experimental observations led to a further subdivision of SMC into Supplementary Motor Areas (SMA) proper, engaged in the control of the muscolo-skeletal apparatus, preSMA covering the same functions but to a lesser extent and the Supplementary Eye Field (SEF) specialized in the control of the eye [Tanji, 1994]. Note, however, that this original distinction has been a subject of a debate as the electrophysiological characterization of cells ’tuning-curves’ in the SMC suggests a relative rather than an absolute difference between subregions, which indicates perhaps a functional gradient rather than a clear dichotomy [Nachev et al., 2008]. Despite the initial framing of its functions in the realm of motor control, the precise contribution of SMC to behavior remains under discussion. Variegated, but not necessarily contrasting evidence suggest that SMC may play a role in volitional execution [Eccles, 1982, Deecke and Kornhuber, 1978, Goldberg, 1985] of internally generated, rather than sensory-driven, movements [Mushiake et al., 1990]. Alternative views suggest its involvement in learning novel conditional stimulus-responses associations [Chen and Wise, 1997, Nakamura et al., 1998].

Importantly for the present discussion, there is a relative consensus about the involvement of SMC in the control of action under cognitive or executive demands [Nachev et al., 2008]. In particular, the SMC is thought to be necessary for proactive switching during ’change of plans’ situations. For example, Rushworth and colleagues [Rushworth and Hadland, 2002] suggested, in a combined fMRI and TMS study, that intact human preSMA is crucial for the inhibition of current automatic motor plans and facilitation of ’deliberate’ motor responses, but does not mediate automatic responses alone. Similarly, [Isoda and Hikosaka, 2007] reported an engagement of neurons in the monkey SEF, where an increased firing rate and artificial stimulation was associated only with gazes triggered by an unpredictable contextual switch (fig. 7-A).

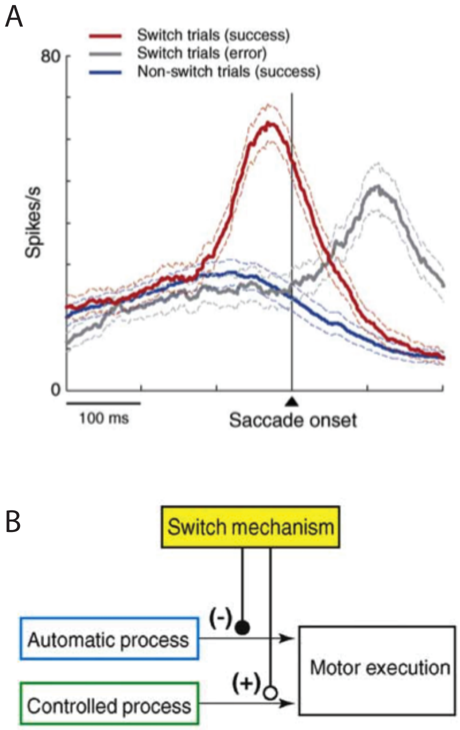


Figure 7: **Switch mechanism in SMA A.**Temporal onset and amplitude of the firing rate in monkey pre-SMA neurons distinguishes between correct and incorrect switch (from [Isoda and Hikosaka, 2007]).**B.**Conceptual model of the switching mechanism possibly found in the medial frontal cortex (from [Hikosaka and Isoda, 2010b]).

Crucially, the temporal onset and magnitude of the response determined the difference between successful and unsuccessful trials. Based on these and a number of other reports, an interpretation of the functional role of the SMC can be therefore bootstrapped from experimental and anatomical evidence, which is, SMC could be a neural controller in charge for switching from automatic to controlled processes. Nevertheless, one of the complexities in the study of this area and its functions is the lack of an established theoretical model that can provide predictions to be tested experimentally [Nachev et al., 2008].

One, and perhaps, the only attempt to define the functional requirements of the supplementary motor complex as a neural controller mediating automatic and deliberate processes comes from [Hikosaka and Isoda, 2010b] who argued that SMC should implement the following functions: (1) detecting a change in context through sensory evidence, (2) suppressing ongoing automatic processes and, most importantly, (3) facilitating the execution of an alternative controlled process (fig. 7-B). In addition, as an extension of the original proposal, we argue that it should also be able to access internal representations (i.e. memory) to inform behavioral choices in the case that environmental cues do not indicate what alternative action should be taken. This last requirement seems of particular relevance in (i.e. social) contexts where norms and rules demand behavioral strategies that are implicit in, for example, the moral code.

0.3.4 Theta oscillations in cognitive control

A natural question following the 'design' requirements postulated by the proposed model of SMC functions is whether evidence for the control signals mediating each of these functions can be found, at the implementation level, in the brain. A number of studies have addressed the purpose to characterize the neural dynamics involved in switching behavior. Importantly, the majority of reports involving human and primate electrophysiology seems to agree on the peculiar role of the oscillatory dynamics in mediating different aspects of switching behavior. In particular, oscillations in the theta range (4-9 Hz) have been greatly implicated in several aspects of cognitive control [Voytek et al., 2015, Sweeney-Reed et al., 2017], leading to the hypothesis that dynamics in this restricted set of frequencies could reflect the neural control signal underlying the coordination of automatic and deliberate processes.

Remarkably, theta phase synchrony between SMC and Parietal and Occipital cortex was found in monkeys who were engaged in a gaze-switching task (similar to [Isoda and Hikosaka, 2007]. Importantly, this synchrony occurred only in the trials which required a switch between contexts [Phillips et al., 2014]. This observation possibly indicates that the fronto-parietal network could constitute the neural pathway through which the detection of a contextual change is conveyed to SMC. In addition, the theta band together with beta band (12-30 Hz) has been widely implicated in the suppression of ongoing motor behavior. For example, a number of EEG studies in human cognitive control report an increase in the power of theta in the mid-frontal electrodes during situations of conflict [Cohen, 2014a]. Here, the increased theta power positively correlates with subjects' response time during an increased demand for cognitive control, suggesting a signature of inhibition over the motor system. An interpretation supported also by a similar relationship with retrospective behavioral changes such as post-error slowing.

On a similar vein, [Chen et al., 2010] observed a comparable role of the theta and beta band in monkey SMA during a countermanding paradigm that only required to retract ongoing arm movements. Here the onset and the peak of the relative change in power were correlated with the greater ability of the animal to inhibit behavior prospectively and retrospectively (i.e. post-error slowing). Note however that retrospective changes are often related to the role of the anterior cingulate cortex [Hikosaka and Isoda, 2010a]. Finally, the inhibitory power that the SMC might exert over the motor system is supported by its anatomical (hyper direct pathway) [Aron and Poldrack, 2006] and functional connectivity with the subthalamic nucleus, a region of the basal ganglia thought to act as a global brake for the motor system [Cavanagh et al., 2011].

In conclusion, there seems to be a convergent empirical evidence that low frequency oscillations could mediate various aspects of deliberate control and that theta frequencies could be involved in synchronizing multiple areas of a distributed network.

0.3.5 Open questions

An increasingly accepted view is that deliberate control is not a localized function in the brain but rather a widely distributed network that comprises a number of functionally specialized sub-systems. These sub-systems are in charge for goal representations, possibly driven by internal states and drives, for memory from previous experience as well as current perceptual evidence and value associated to it. For example, there it has been proposed that the frontal cortex could be orchestrating goals, while the hippocampus and the temporal lobe could hold memories that allow planning based on previous experience. However, it remains unclear how this wide network is integrated into a coherent behavioral response.

The second aspect related to the neurophysiology of deliberate control pertains to the role of low frequency oscillations. In particular, theta oscillations could mediate distinct aspects of deliberation as it has been found during cognitive control tasks [Cavanagh et al., 2010, Helfrich and Knight, 2016], but also, more generally, in a wide range of cognitive functions including memory and visual attention [Lisman and Jensen, 2013, Szczepanski et al., 2014]. However, it remains unclear what is the role of theta oscillations and the function of distinct features of the neural signal, such as amplitude and phase, in orchestrating goal oriented behavior. Indeed, increases in amplitude have often been reported during cognitive or deliberation task as a neural correlate of conflict. Phase, in turn, seem to have a long-range synchronization mechanism. It remains unclear, however, whether phase dynamics are solely a mean to orchestrate the different sub-systems involved in goal switching or perhaps they could also constitute a control signal that drive behavior. Answering this questions would contribute to elucidate not just the oscillatory dynamics involved in this type of behavior but it would also address a broader, and perhaps pressing issue in neuroscience regarding the functional role of oscillations in the brain.

0.4 Thesis outline

0.4.1 Part 1

In the first contribution of this dissertation, we revisit the concepts of automatic and deliberate processes in the light of the classical insights on animal learning. We do this from the perspective of biological control modeling through the implementation of a brain-based control architecture. In particular, consistent with the tradition of the experimental psychology, we choose a foraging task where a mobile robot has to navigate an open maze in search for distinct resources that need to be hoarded in the nest location. The aim of this study is to provide an explicit mapping of automatic and deliberate processes to the neural substrate of distinct regions of the brain by bridging between the information processing capabilities of individual modules and the global objective of controlling behavior through their integration. We demonstrate how reflexive, automatic and deliberate behaviors are the result of the interaction of individually distributed sub-systems that progressively develop associations and internal representations in a bottom-up fashion through the situatedness and embodiment of the system.

0.4.2 Part 2

In part two of this dissertation, we focus on the automatic control properties of the cerebellum. We do this from a computational and control perspective by modeling the acquisition of anticipatory postural adjustments in a robot. Starting from a standard cerebellar control scheme based on FEL, we provide an initial explanation for this phenomenon that relies on the integration of multiple, parallel adaptive responses triggered by predictive cues. We observe, however, that this control scheme lacks the robustness to cope with uncertainty. By looking at the anatomical organization of the cortex and the cerebellum, we introduce the notion of hierarchy and later translate this biological insight into a control architecture. We further propose a new theoretical framework where the FEL architecture can be recast into a hierarchical scheme that reflects the causal structure of perceptual events in the environment. This novel approach can achieve anticipation while solving the control problems of violation and generalization in a simulated biologically-plausible postural task, and it exhibits a performance consistent with behavioral literature. This theoretical framework has implications for cerebellar theory as it proposes that hierarchical descending predictions and ascending prediction-errors can be implemented in cortico-cerebellar loops that only modulate behavior at the last stage of the control chain, as also proposed by normative accounts such as the active inference.

0.4.3 Part 3

In part 3 of this dissertation, we focus on deliberate control of action within the medial frontal cortex. We take a neurophysiology approach with the objective to explore the oscillatory dynamics that underlie the switch from automatic to deliberate actions in the human brain (iEEG). In particular, we sought to determine the role of theta oscillations in controlling switch behavior as well as the network interactions of the medial frontal cortex with other areas involved in deliberate control. Our results suggest that Supplementary Motor Areas (SMA) facilitate switching behavior through a phase code. In particular, we find that the phase alignment of a small subset of frequencies in the theta range is predictive of performance (Reaction Times, RT) at the trial level, with theta-gamma phase amplitude coupling being higher in faster trials. In addition, we find that the in switch trials SMA synchronizes with the temporal lobe in a behaviorally dependent manner. These results contribute to elucidate the role of theta oscillations in controlling deliberate actions and support the involvement of a distributed brain network engaged in cognitive control.

Part I

Chapter 1

A BRAIN ARCHITECTURE FOR AN ARTIFICIAL FORAGER

In this chapter, we propose an integrated brain architecture for an artificial forager that reflects the anatomical organization of the rodent brain with the goal to explore the internal dynamics of individual brain modules and their contribution to the control of behavior and, more generally, to the global fitness of the agent. We do this in the context of a foraging task, where an artificial agent has to navigate a complex environment and collect different types of resources while avoiding obstacles. Here, a number of computational models reflecting the core computation of multiple brain areas, such as the cerebellum, hippocampus, and frontal cortex provide multiple levels of control from reflex-like behavior to automatic adaptive control to deliberate decision-making. We show that an incremental activation of hierarchical levels of control produces incrementally complex foraging strategies, from random navigation to stimulus-response associations and finally to planning with an overall increase in the optimality of the behavioral strategy.

This chapter is based on: *Maffei, G., Santos-Pata, D., Marcos, E., Sanchez-Fibla, M., & Verschure, P. (2015). An embodied biologically* *constrained model of foraging: from classical and operant conditioning to adaptive real-world behavior in DAC-X. Neural Networks, 72, 88-108.*

1.1 Introduction

Animals act in order to survive, reproduce and enhance their fitness in general. It has been proposed that the functions of the brain behind this performance can be decomposed into 5 top-level objectives called: ”how”, ”why”, ”what”, ”where” and ”when” or the H4W problem [Verschure, 2012]. A paradigmatic example of H4W is the foraging task where an animal must both explore and exploit its environment in order to survive. This form of complex behavior includes: to learn where and when to look for resources, what to look for, where and when to return to the home base, how to avoid obstacles and how to act in order to satisfy internal needs, see [Clayton and Dickinson, 1998]. Animal foraging and in particular hoarding, where an animal stores resources for later retrieval, has greatly contributed to the understanding of animal behavior from a neurobiological perspective in ecologically valid conditions. For example, the selection of need-relevant resource types for hoarding has been shown to be dependent on motivational systems defined by the Hypothalamus and associated Brainstem nuclei [Keen-Rhinehart et al., 2010]. The same study has shown that hamsters tend to apply an energy efficient hoarding strategy (see also [Cabanac and Swiergiel, 1989, Lea and Tarpy, 1986]) consistent with optimality accounts of foraging [Charnov, 1976]. In addition, hoarding tasks are greatly used in the experimental study of spatial cognition and its neural substrate, in particular the Entorhinal cortex and Hippocampus. For example, hoarding behavior can be realized using either an egocentric strategy combining local cues and/or landmarks through associative learning with actions, as we might find in the substrate of classical conditioning, or an allocentric one that relies on global map-like representations found in the Hippocampus [Tolman, , van der Meer et al., 2012, Gould et al., 2010, Smulders et al., 2010, Buzsáki and Moser, 2013]. Despite insights into the role of single neural populations in hoarding, the principles governing the interaction among the multitude of processes underlying complex goal-oriented behavior, such as foraging, are poorly understood. This is partially due to the experimental intractability of the system-level and multi-scale organization of the neural principles underlying most behaviors. It is exactly for this reason that a computational exploration can assist in advancing hypotheses on the system level properties of the whole brain as it engages with the world through the body.

Several aspects of complex animal behavior have been previously described within the context of the learning paradigms of classical and operant conditioning [Thorndike, 1927, Hull, , Skinner, 1948, Abarca et al., 1985, Agetsuma, 1998, Domjan, 2004, Fantino and Abarca, 2010]. In classical conditioning, an animal learns to relate an initially neutral stimulus with an innate reflex that comprises a predefined stimulus-response relation. Through experience the initially neutral stimulus will substitute the reflex-triggering stimulus in driving the reflex in a predictive manner. Examples can be found in anticipatory salivation and mastication preceding food intake [Pavlov and Anrep, 2003, Woods and Ramsay, 2000, Woods, 1991], courtship behavior triggered by the presence of a possible partner [Zamble et al., 1985] and avoidance responses anticipating noxious stimuli as in eyeblink conditioning [Gormezano et al.,] (see [Domjan, 2004] for review). In operant conditioning, the contents of what is learned are the consequences of action and their relationship to the stimulus context. The outcome of this learning is determined by the context and the goal to be achieved [Thorndike, 1927]. Instrumental tasks, where an animal learns to increase the probability of obtaining a food reward by performing a specific action (e.g.. lever pressing) in a specific situation (context) have been often used to study foraging [Abarca et al., 1985, Jurado-Parras et al., 2013]. At present, however, there have been no attempts to reconcile the multitude of neural mechanisms underlying foraging with the behavioral modulation identified by the two conditioning paradigms in a biologically grounded cognitive model. Here we attempt to overcome this limitation by proposing a model that integrates the computational functions of several brain areas thought to play a role in the two learning paradigms of classical and operant conditioning while accounting for foraging.

We realize our model from the perspective of the Distributed Adaptive Control (DAC) theory of mind and brain that interprets these two paradigms in terms of core functional subsystems underlying adaptive behavior, organized into a four-layered cognitive architecture: somatic, reactive, adaptive, contextual [Verschure and Voegtlin, 1998] (see [Verschure, 2012] for review and [Verschure et al., 2014b] for a detailed mapping to the neural system underlying goal-oriented behavior). According to DAC the somatic layer defines the fundamental interface between the embodied agent and its environment, including the needs that must be satisfied in order to assure physical integrity and survival. It captures physical properties, such as body morphology, sensorium and effectors and accounts for the role these play in the generation of behavior as for instance expressed in the concept of morphological computation [Pfeifer et al., 2007]. The reactive layer describes innate behavioral systems comprising reflexes and low-level stereotyped behavioral patterns, such as those driven by fight or flight instincts and dedicated triggering stimuli, as famously described in the hierarchy of sexual behaviors of the three-spined stickleback fish by Tinbergen [Tinbergen,]. These behavioral subsystems are genetically predefined as sensory-affect-motor mappings, provide a first level of internal processing driving somatic responses (i.e. reflexes) and account for the putative computational functions of Brainstem nuclei such as the Pons [Bracha et al., 1991], Central Grey [Panksepp and Biven, 2012] and the Superior Colliculus [Meredith and Stein, 1983]. Within the reactive layer, behavior regulation follows homeostatic and allostatic principles [Fibla et al., 2010] driven by external stimuli or somatic states, defining a first step towards the generation of goals coherently with the integrative functions ascribed to the Hypothalamus [Blouet and Schwartz, 2010, Sutcliffe and de Lecea, 2002, Volkow et al., 2011, Pfaff, 1999]. The reactive layer captures the so called ”emotional operating system” proposed by Panksepp that identifies seeking, fear, rage, panic, lust, care and play as the primary behavioral subsystems of the mammalian brain [Panksepp and Biven, 2012]. Alternatively, this layer also implements the circuits underlying the Unconditioned Stimuli and Responses of classical conditioning [LeDoux, 2012]. It is important to emphasize that the reactive layer provides, through its control over the main neuromodulatory systems, global regulatory signals that strongly affect higher levels of organization [Swanson, 2012]. As such, it provides the foundation on which the content of memory systems is bootstrapped. The adaptive layer of DAC captures perceptual and behavioral learning systems such as the stimulus-stimulus and stimulus-response associations studied in classical conditioning [Pavlov and Anrep, 2003, Gormezano et al., , Domjan, 2004]. At this level, the acquired value of sensory inputs is shaped by experience and leads to the anticipation of conditioned responses through the predictive control of perception and action. This frees the agent from the restricted envelope of phylogenetically defined reflexive systems tuned to evolutionarily invariant trends of survival and allows it to adaptively shape its actions to a priori unknown states of its environment. The learning mechanisms of the adaptive layer are composed of the Amygdala [LeDoux, 2012], primary sensory areas of the Neo-cortex [Weinberger, 2004] and the Cerebellum [Medina et al., 2002] as described in the so called two-phase model of classical conditioning [Inderbitzin et al., 2010, Konorski and Miller, 1937, Sánchez-Montañés et al., 2000, Clark and Squire, 1998]. In particular, the Cerebellum is at the center of the short-timed sensorimotor learning component of classical conditioning and provides key interfaces to the reactive layer by being recurrently coupled to its core nuclei. The adaptive layer also accounts for internal mechanisms of reward and motivation, such as those attributed to the Striatum [Pennartz et al., 2011] and the ventral tegmental area [Lockie and Andrews, 2013], considered to be a crucial component of the brain rewarding system that drives the motivation for action in operant conditioning [Luo et al., 2011]. Finally, the contextual layer describes the goal oriented decision-making abilities of the brain, built on sequential memory systems. Here, higher-level representations, such as sensory percepts, behavioral policies, encoding of space and goal definitions provide the foundation for learning in a broad temporal window dependent on the actions of the agent as observed in operant conditioning, where conditioned responses are actively generated, situation dependent and task dependent [Jurado-Parras et al., 2013, Luo et al., 2011]. The contextual layer comprises short- and long-term memory systems that allow the formation of goal-oriented sequences of sensorimotor representations and that is closely tied to our understanding of the Hippocampus [Hafting et al., 2005, Lisman, 2005, O'Keefe, 1976], medial Prefrontal [Miller and Cohen, 2001] and Premotor cortical regions [Marcos, Encarni Verschure, 2013] and the dynamics regulating their interaction [Blumenfeld and Ranganath, 2007].

On one hand, the general principles of DAC and its layered organization of behavior have been validated in several previous studies using both mobile and humanoid robots [Verschure et al., 1995, Verschure and Voegtlin, 1998, Verschure et al., 2003b, Duff et al., 2011, Duff and Verschure, 2010, Verschure, 2012]. These studies however used abstract computational models that did not take into account a functional mapping to specific brain areas. On the other, specific hypotheses on individual structures have been recently validated within the context of DAC, by adding biological constraints to its theoretical principles and aiming at a more faithful mapping to the computational processes found in the brain. These include the Cerebellum [Herreros et al., 2013b, HofstoÈtter et al., 2002], the entorhino-hippocampal formation [Guanella et al., 2007, Rennó-Costa et al., 2010] and the Prefrontal and Premotor cortex [Marcos, Encarni Verschure, 2013]. To date however, these models have only been validated in isolation against specific data sets and have not yet been brought together in a single integrated system generating behavior from the interaction between its multiple layers. To bridge this gap we present a cognitive architecture, called DAC-X, that unifies the theoretical principles of DAC with biologically constrained models of several areas of the mammalian brain, working synergistically, in real-time, to control a robotic agent engaged in a hoarding task. We embed in DAC-X the computational features characterizing the neural substrate thought to be responsible for classical and operant conditioning with the assumption that they could constitute a sufficient computational architecture able to account for complex behaviors, such as foraging, and ultimately solve the H4W problem. In particular, we implement a Hypothalamus based behavioral regulation system called Allostatic Control [Fibla et al., 2010] integrated in the reactive layer to manage the drives of the agent and their associated behavior systems, a model of the Cerebellum used to learn adaptive motor responses to cues and to perform predictive obstacle avoidance, a model of the Hippocampus used to build the cognitive map of the agent and solve its self-localization problem (SLAM) and providing the foundation for planning capabilities in the contextual layer, and finally a Prefrontal based decision-making model managing, at the contextual level, the selection of goals and plans. To benchmark our system we created a setup composed of a robotic agent and a foraging arena [Dasgupta et al., 2014, Verschure and Voegtlin, 1998, Verschure et al., 1995, Prescott et al., 2006]. The goal of the task is to forage two types of motivationally relevant resources within the environment and bring them to the home location by exploiting the main perceptual and cognitive abilities found in animal hoarding, such as exploration, learning, conflict resolution, decision-making, navigation and associative learning. We will show that the interaction between the layers of the DAC-X architecture, individually learning through behavioral feedback, can capture the intrinsic dynamics and the behavioral outcomes described in both classical and operant conditioning paradigms and can efficiently control an agent performing a foraging task within a complex environment. Our results indicate that optimization of individual error functions driving different subsystems can positively affect the efficacy of the hoarding strategy pursued by the agent, described in terms of cost-reward relationship. We discuss our results in the light of behavioral and neuroscientific evidence from different fields of psychology and ethology and describe how our system can account for decision-making, navigation, learning and other cognitive aspects found in foraging and hoarding, ultimately addressing the H4W problem. We also describe how DAC-X is consistent with behavioral and learning dynamics found in the paradigms of classical and operant conditioning. Finally we discuss our modeling approach towards mind and brain in comparison to previous implementations of the DAC theoretical framework and other recent computational approaches devoted to the understanding of the principles and the neural substrate driving goal-oriented behavior.

1.2 Methods

In order to benchmark the performance of DAC-X, we devised a robot-hoarding task within an open arena with obstacles, that includes several behavioral, perceptual and cognitive aspects found in animal foraging and hoarding. We tested the behavior of the agent in the execution of the task and analyzed the evolution of its performance as a function of its experience and its underlying computational principles.

1.2.1 Setup

We used a custom built robot platform (diameter: 7cm) to perform the task (fig. 1.1). It comprises a set of 5 infra-red proximity sensors (Sharp 2Y0A21; range: 0-20 cm,) evenly distributed around its base, two directional active wheels (Springrc SM-S4303R), a video camera (Pixy 1.3, Charmed Labs; resolution: 640x480 pixels) for object detection, a digital compass (Xsens, MTw Development Kit Lite), and a one-DOF active robot gripper (custom made) for object carrying (fig. 1.1-B,C). An onboard microcontroller (Ardunio Uno) samples all sensor signals and actuator commands at 50 Hz. Computations were performed on a desktop computer (MacBook Pro, Apple) and robot-computer communication was established via Bluetooth. The foraging arena is a surface of size 100x60 cm [Fibla et al., 2010] comprising: a home location characterized by a yellow patch, a set of objects representing resources of two types (blue objects represent water, red objects represent food), a set of rigid walls acting as obstacles, and finally a set of green patches on the floor used as cues near collectable resources (fig. 1.1-A). A video tracking system (Reactivision tracking software) mounted underneath the translucent table top was used to record the robot's and objects' positions during the task at 25 Hz.

1.2.2 Experimental Paradigm

Task

In order to evaluate the behavioral performance of DAC-X in the context of foraging behavior, we defined an experimental protocol based on a real-world hoarding task within an open-maze arena. In this context the agent had to display all features known from rodent hoarding behavior [Pravosudov and Smulders, 2010], such as taking decisions based on internal states [Keen-Rhinehart et al., 2010], rely on associative learning of landmarks with locations of resources and spatial memory [Gould et al., 2010, Smulders et al., 2010], in order to optimize the resource gathering process [Cabanac and Swiergiel, 1989, Lea and Tarpy, 1986]. Consistent with similar animal [Jurado-Parras et al., 2013, Kim et al., 2013] and robotic tasks [Montague et al., 1995, Verschure et al., 1995, Verschure and Voegtlin, 1998, Prescott et al., 2006, Dasgupta et al., 2014] the agent was required to explore a patched environment in search of resources, starting from an initial home location. The resources were suitable to reduce the drive provided by two simulated internal needs (e.g. hunger or thirst). Once the correct drive-reducing resource was found, the agent had to grasp it and bring it to the home location in order to obtain the specific reward. Such a sequence, describing the course of one trial, from the home location departure to the hoarded resource, was repeated along the duration of one experimental session until the resources (12 items in total) were exhausted. At the beginning of every session the instantiation of DAC-X (section 2.3) was reinitialized with all memory systems erased, preventing the reuse of any knowledge of the environment acquired in previous sessions. The main goal of the agent was to increase the efficiency of the hoarding process over trials, where efficiency was defined as the relationship between the obtained reward value associated to a collected item and the energy spent to obtain it. To do this, the agent had to make a decision on what type of resource to fetch at every trial, to efficiently navigate, avoid obstacles and walls, and acquire and exploit both specific (cue and action associations) and contextual (space and trajectories) knowledge.

Experiment

In order to evaluate DAC-X we ran 9 experimental sessions of 12 trials each with the aim to evaluate the behavior of the robot and its efficiency in hoarding resources. To do this we ran the system with a full implementation of the control architecture (described in section 2.3) and we analyzed the performance over different stages of learning together with the underlying neuronal dynamics.

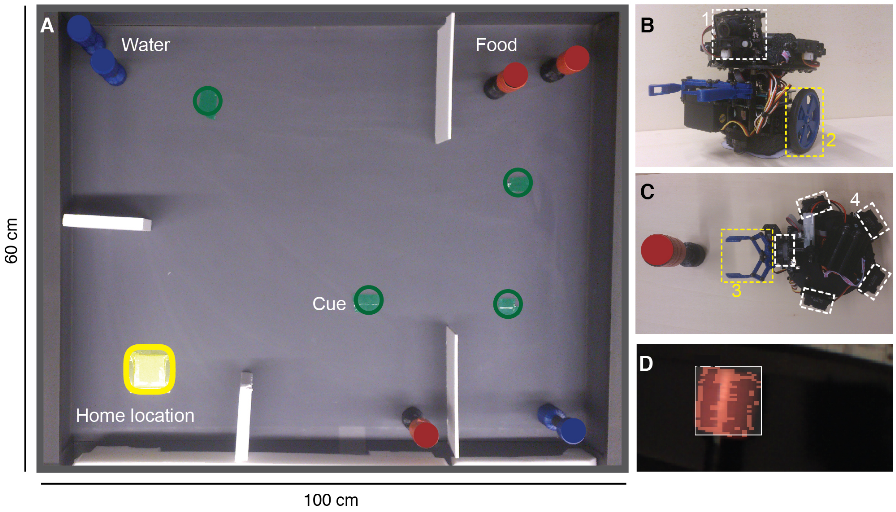


Figure 1.1: **Robotic setup. A.** Foraging arena. Red represents food resources. Blue represents water resources. Yellow represents the home location (patch on the floor). Green represents visual cues for sensory motor associations. **B,C.** Robotic agent. 1: camera. 2: active wheels. 3: gripper. 4: proximity sensors. **D.** Visual field of the robot facing a food object identified using a color blob detection algorithm.

In particular, we analyzed behavior in terms of navigation, quantifying trajectory lengths and occupancy patterns over the course of the session. We looked at efficiency in terms of the relationship between cost and reward as the main measure of performance, where the former captures the total motor activity (see section 2.3.4) produced in exploratory behaviors and item collection, while the latter captures the reward value associated to each item and discounted over the time of a trial (see section 2.3.3.2). We also provided an analysis of the dynamics that regulate individual processes of the main modules of DAC-X and quantify their evolution during the session. In addition, we ran 3 sets of 5 experimental sessions each with the aim to assess the individual contribution of each layer of the architecture over the progression of learning and performance. To do this we progressively enabled a layer of control at every set, starting from reactive (R), to reactive-adaptive (R+A), to reactive-adaptive-contextual (R+A+C) and compared evolution of behavior and hoarding efficiency for these three conditions over trials.

1.2.3 Control Architecture

The behavior of the agent was generated by a cognitive architecture based on the Distributed Adaptive Control (DAC) theory of mind brain and behavior (Verschure & al., 2003; Verschure, 2012). According to DAC, the behavior of organisms can be understood in terms of the interaction between multiple layers of control. Here, each layer implements complementary computational strategies, on various degrees of biological validity, mapping sensory inputs to internal states and from states to action. As reported in detail below, we mixed controllers that only emulate putative behavioral functions of a specific brain structure with computational models constrained by biological validity (fig. 1.2). For each of the models, we will explain in detail their physiological and anatomical grounding and prior work on the basis of which we justify their biological validity.

The somatic layer of DAC-X represents the physical properties of the agent, such as body morphology, sensors and actuators. The reactive layer of control represents the lowest computational level of the architecture and reflexively maps sensory states into action as described above. Here, we used a set of feedback controllers that aim to approximate the role of reactive Brainstem nuclei. In particular we modeled a controller responsible for avoidance reflexes, that resembles in its function the reactive pathway found in the Trigeminal Nucleus and its motor output (via Reticular formation and Abducens motor nuclei; [Bracha et al., 1991]). In addition we implemented a controller responsible for visuomotor reflexes used to issue orienting responses towards a visual target which approximates the behavioral output of the Superior Colliculus [Goodale et al., 1975, Meredith and Stein, 1983]. Finally, we created a set of leaky integrators computing the somatic states of the agent, such as internal needs and low-level drives for action, with the aim of approximating the monitoring role attributed to the Hypothalamus [Volkow et al., 2011, Fibla et al., 2010].

The adaptive layer of control is responsible for associative sensory-motor learning, encoding of motivation and reward, and contributes to action selection. Multiple instances of a biologically plausible model of the cerebellar microcircuit learned to associate initially neutral sensory cues with adaptive responses under the supervision of reactive modules [Herreros et al., 2013b, Inderbitzin et al., 2010]. A module mimicking the putative computation performed by the Ventral Tegmental Area served as a transformation step from low-level internal states to motivation for action, laying the foundation for the definition of high-level goals [Lockie and Andrews, 2013, Verschure et al., 2014b]. An abstract implementation of the Basal Ganglia performed action selection for behavioral plans defined at higher levels of control. In addition, the Basal Ganglia module also delivers reward signals to hypothalamic nuclei (the core structure of the reactive layer) as observed in [Pennartz et al., 2011]. The contextual layer provides the memory systems underlying spatial cognition, decision-making and planning. It comprises a biologically constrained model of the Hippocampus [de Almeida et al., 2009, Rennó-Costa et al., 2010] that acquires an internal representation of the environment underlying self-localization, mapping, navigation and spatial planning. Finally a model of the Prefrontal Cortex implements a biologically valid computational mechanism for decision-making [Marcos, Encarni Verschure, 2013] and a more abstract one mimicking the functions of long-term memory for goal-dependent information storage [Duff et al., 2011].

Below we describe in detail each layer of DAC-X and its main computational features.

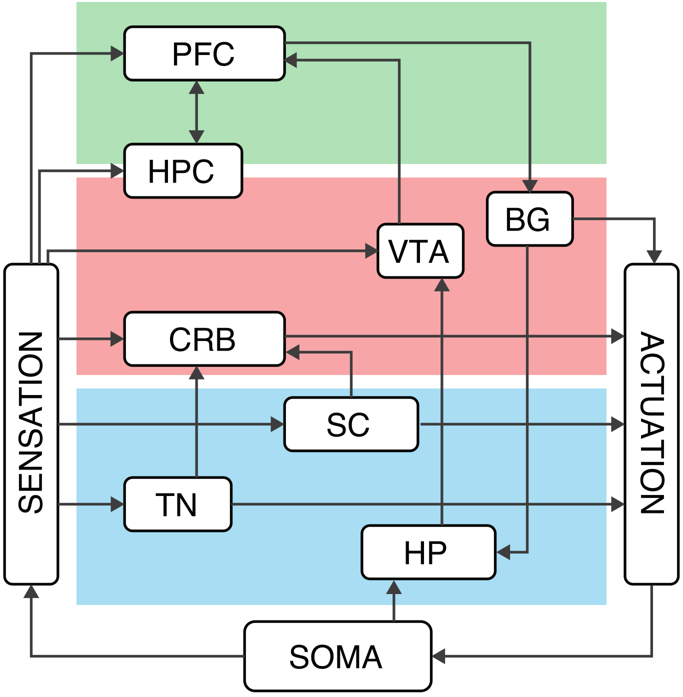


Figure 1.2: **DAC control architecture.** From bottom to top: the soma represents the physical properties of the agent. The reactive layer (blue) of control is in charge of mapping sensory states into prewired action. This layer includes computational elements mimicking the reflex-like sensory-motor transformations performed by some Brainstem components such as Trigeminal Nucleus (’TN’) and the Superior Colliculus (’SC’). The module mimicking the functions of the Hypothalamus (’HT’) encodes somatic states as internal needs. The adaptive layer (red) of control is in charge for associative sensory-motor learning, encoding of motivation and reward and action selection: this layer includes multiple models of the cerebellar microcircuits (’CRB’), a model approximating the functions of the Ventral Tegmental Area (’VTA’) and one performing some of the computation attributed to the Basal Ganglia (’BG’). The contextual layer (green) is in charge for spatial representation, decision-making and planning: this layer includes a computational model of the Hippocampus (’HPC’) and

a computational model of the Prefrontal Cortex ('PFC'), together with a long-term memory structure for goal-place couplets storage.

Somatic Layer

The somatic layer represents the body of the agent as an interface with the world (fig. 1.1-B,C). It describes low-level processes such as exosensing (perception of the environment) and actuation. Here the relevant somatic input signals were computed by the hardware components of the agent, such as the proximity sensors, used to provide information on the relative position of obstacles and walls. In addition we performed color-blob filtering (colors: red, green, blue and yellow) on the source image provided by the on-board camera to extract the coordinates of objects and relevant landmarks within the visual field of the agent (fig. 1.1-D). Finally the somatic outputs are computed as the total motor signal conveyed by the control architecture to the motor wheels for spatial navigation and to the gripper actuator for object recollection.

Reactive Layer

Allostatic control in the Hypothalamus. The simulated needs of the agent encoded in its hypothalamic states represent glucose and hydration levels, driving exploratory, consummatory, and security behavioral subsystems [Blouet and Schwartz, 2010]. Similar to the work of [Fibla et al., 2010] the state of each subsystem, $s(t)$, was modeled as a leaky integrator with a dynamic decay rate:

$$s(t) = s(t-1) - kAs(t-1) + v_r(t_{tr})$$

(1)

where v_r represents the reward value associated to a specific state delivered to each subsystems by the adaptive layer (see equation 14), A is the decay rate (here set to 0.001) that is continuously modulated by a gain k encoding two alternative elements: the total motor activity at every time step (normalized between 0.1 and 1) for glucose and hydration values, or the Euclidean distance of the robot from the home location (normalized between 0 and 1), for security levels (Euclidean distance computed from tracking system coordinates, but see [Bernardet et al., 2008] for a biologically grounded implementation). As in the notion of allostatic control [Fibla et al., 2010], hypothalamic states can trigger basic predefined behaviors such as freezing or reactive navigation (see equation 6), useful to reset the homeostatic levels of each variable within the same layer. Hypothalamic values are also used to set low-level resource seeking, driving the attentional system.

Reactive attentional mechanisms of the Superior Colliculus. A basic attentional system used to orient the agent towards a visual target takes into account the high-level computational features found in the Superior Colliculus (SC), where the sensory-to-motor gradient found from the dorsal to the ventral part of the SC is believed to map retinotopically organized visual inputs into reflexive actions [Goodale et al., 1975]. Analogously, we used a proportional feedback controller that maps the position of a target detected in the visual field into a motor response, t_{SC}, which centers the fovea of the agent on the target location, according to:

$$t_{SC} = K(x_{obj} - x_{fov})$$

(2)

where x_{obj} is the horizontal position of the centroid of an object in the visual field, computed as a color blob by camera image processing in the somatic layer (section 2.3.1 and fig. 1.1-D) and normalized between 0 (extreme left) and 1 (extreme right). x_{fov} represents the normalized horizontal position (= 0.5) of the center of the visual field. Finally K represents a constant gain, that is set to 3.

The resulting command generates an agonist-antagonist turning action directed to the left, w_{left}, and right wheel, w_{right}, such that:

$$w_{left} = m_r + t_{SC}$$

$$w_{right} = m_r - t_{SC}$$

(3)

where $m_r (= 3)$ represents a constant forward motor command of 2 cm/sec reflecting a constant exploration drive.

Brain stem avoidance reflexes The avoidance reflexes mimic the computational functions attributed to the Brainstem and in particular to the Trigeminal Nucleus (TN) that directly maps noxious sensory stimuli into avoidance motor responses by driving the spinal cord and the musculoskeletal system [Bracha et al., 1991]. Here we implemented a reactive feedback controller, responsible for collision avoidance, which proportionally maps the proximity signal detected by a somatic sensor p_j, into an avoidance reflex, t_{TN} , according to:

$$t_{TN} = (K_{p1}p_1 + K_{p2}p_2 + K_{p3}p_3) - (K_{p4}p_4 + K_{p5}p_5)$$

(4)

where j denotes the sensor index from extreme left (1), to center (3), to extreme right (5) and K_{pj} an asymmetrical scaling factor. Each scaled component is linearly summed and mapped to the left, w_{left} , and right, w_{right} , wheel to perform collision avoidance following:

$$w_{left} = m_r + t_{TN}$$
$$w_{right} = m_r - t_{TN}$$

(5)

Altogether the reactive layer in the present system is in charge for monitoring the internal state of the agent, providing a drive for action to the above layers, as well as implementing a set of reflex-like sensory-motor mappings, driving reactive navigation strategies. Reactive layer motor control is thus defined as the summation of exploration, orientation and avoidance:

$$w_{left} = m_r + (t_{TN} + t_{SC})$$
$$w_{right} = m_r - (t_{TN} + t_{SC})$$

(6)

where the linear sum of the commands provided by the obstacle avoidance system, t_{TN} , and the orienting system, t_{SC} , is used to control the agonist-antagonist turning gains applied to the motors, while m_r controls the constant forward gain. Finally, the reactive layer provides teaching signals to the adaptive layer together with efferent copies of reactive motor commands (see section 2.3.3).

Adaptive Layer

Adaptive motor control in the Cerebellum. In order to achieve a computational strategy useful to acquire adaptive motor responses we implemented a biologically valid model of multiple parallel cerebellar microcircuits. Each microcircuit is in charge for acquiring the predictive value of initially neutral sensory cues, such as color patches or obstacle proximity values and associating it with appropriate motor responses.The computational mechanism is based on an analysis-synthesis adaptive filter implementation mimicking the learning properties of the cerebellar microcircuit [Herreros et al., 2013b, Maffei et al., 2014]. Note that, unlike previous implementations [Duff and Verschure, 2010], the perceptual learning component and the sensory-motor learning component were merged into the same algorithm.

According to the present setup, in each microcircuit an acquired output signal is obtained by transforming the input signal into a target signal under the supervision of an error signal. To achieve this, the input y_t to an adaptive module is expanded in a set of bases (20 basis functions per microcircuit in the current setup). Each basis results from a fast excitatory component, exc, subtracted by a slow inhibitory one , inh. Each component is obtained as a double convolution with two exponentials in such a way that the response of each basis to a unitary pulse resembles an alpha function. The time constants governing the rise and the decay of each basis are randomly drawn from two flat distributions (a fast time constant, ranging from 10 to 60 ms and a slow one, ranging from 60 to 1000 ms). The value obtained after the two convolutions is then thresholded and scaled for each basis. The output $exc_j(t)$ of an individual excitatory component of a single basis is obtained as:

$$exc_j^r = \gamma_j^r exc_j^r(t-1) + y(t-1)$$
$$exc_j^d = \gamma_j^d exc_j^d(t-1) + exc_j^r$$
$$exc_j = \sigma_j[exc_j^d(t-1) - h_j^{exc}]^+$$

(7)

where j is the index of a particular basis, exc^r and exc^d perform the convolution imposing the temporal dynamics of rise and decay of the output function. γ^r and γ^d represent persistence factors for rise and decay imposing appropriate exponential decay. Importantly, non-linearity of the response is achieved by applying a threshold, h_j^{exc}, to the output signal which is finally scaled by the factor σ_j, where $[x]^+ = max(0, x)$. This threshold factor, drawn from a random distribution within a range of appropriate values, is used to modulate the magnitude and temporal properties of the final response. For each basis, an inhibitory component, inh, is computed in the same way as the excitatory component, but using larger time constants. The final value of the basis function, p_j, is computed by integrating excitatory and inhibitory components such as:

$$p_j(t) = [exc_j(t) - inh_j(t)]^+$$

(8)

The output of a single adaptive module is obtained as a weighted linear combination of its components:

$$z(t) = [p(t)^T w(t)]$$

(9)

where $p(t)$ represents the vector of basis at every time step and $w(t)$ represents the vector of weights. The weights are updated according to the following learning rule:

$$\Delta w_j(t) = \beta e(t) p_j(t-\delta)$$

(10)

where β (set to 10 in the current setup) is the learning rate and $e(t)$ represents the error value at every time step. δ represents the sensory feedback delay set to 800 ms. Finally, the error signal is computed as the difference between a given reactive motor command, m_r, and the weighted output of the Cerebellum:

$$e(t) = m_r(t) - K_{noi} z(t - \delta)$$

(11)

where K_{noi} is a constant gain representing the cerebellar nucleo-olivary inhibitory gain used to balance the reactive and adaptive contributions to the final motor response. Here we set it to 0 so that the acquired response will entirely depend on the external sensory feedback error (see [Herreros et al., 2013b] for details).

Coherently with anatomical evidence, we modeled the interaction between the Brainstem and Cerebellum by conveying an efferent copy of the output of the reactive controllers to individual instances of the cerebellar microcircuit, for a total of seven parallel microcircuits. In particular, we account for bidirectional connections with Superior Colliculus [Kawamura et al., 1982, May, 2006]. In our system, such interaction drives the association of initially neutral visual cues (such as a color patch on the floor, y_t in equation 7) with the reactive responses provided by the SC reactive module when a target object is detected within the visual field (t_{SC} in equation 2). After learning, the resulting adaptive output is an anticipated steering action, triggered by a predictive sensory cue, toward a target resource (i.e. predictive modulation of reactive attentional mechanisms).

Similarly, we model bidirectional connections with the reactive controller accounting for avoidance reflexes found in the Trigeminal Nucleus [Bracha et al., 1991] used in the acquisition of anticipatory adaptive responses in collision avoidance actions, similarly to [Herreros et al., 2013b]. We instantiate one microcircuit for each sensor and provide the proximity signal as input (y_t in equation 7), whereas, the avoidance reflex output by the above described reactive controller is provided as a teaching signal (t_{TN} in equation 4). Such association allows, after learning, more efficient collision avoidance. The total adaptive motor , t_A, output of multiple cerebellar microcircuits is computed as:

$$t_A(t) = \sum_{j=1}^{N} z_j(t)$$

(12) where z_j is the output of a given microcircuit at index j.

Motivation, reward and action selection in the Ventral Tegmental Area and the Basal Ganglia. A second function modeled within the adaptive layer can be found in the motivational system that describes, at a high level, the putative computational role of the Ventral Tegmental Area (VTA) and the system that computes reward value, mimicking the role attributed to the Striatum (STR). As suggested in [Verschure et al., 2014b] and observed by [Lockie and Andrews, 2013], the information processing performed by VTA is thought to drive the motivation to act, by encoding low-level internal states and projecting a modulatory signal towards higher level areas responsible for planning and decision-making. In the current implementation a module approximating the computational role of VTA was used as an intermediate step between the hypothalamic module that encodes somatic needs and the high level decision-making system sitting of the contextual layer, responsible for internal conflict resolution among competing motivational drives. In particular, somatic states encoded by the hypothalamic module, namely glucose level, hydration levels and security levels are transformed into motivational drives according to the following equation:

$$q_s(t) = s_{max} - s(t)$$

(13)

where s represents a current internal state (see equation 1) normalized between 0 and 100, while s_{max} represents its maximum value (= 100). Our model of the Striatum computes the value of a reward used to satisfy internal needs, on the basis of two main properties: first, we implement the association of specific reward value to individual resources in the environment [Pennartz et al., 2011]. Second, we take into account the discounting effect of time on reward value as proposed by [Kobayashi and Schultz, 2008]. The first aspect was modeled as a reward value delivered when a specific object is found and brought back to the home location. Such values were predefined within the architecture as 50 units reward for each food resource, 35 units reward for water resource and 0.001 units reward for every time step spent into the home location for security. Reward values for food and water at total elapsed time from the trial start t_{tr}, v_r (t_{tr}), were discounted as a parabolic function of time at the end of every trial [Shadmehr et al., 2010a] such that:

$$v_r(t_{tr}) = (r(t_{tr}))/(1 + \beta t_{tr})$$

(14)

where r represents the reward associated to a single collected resource and β represents the constant temporal discount rate (here set to 0.0001) .Each final reward value is projected to the allostatic module and integrated in the computation of the internal states of the current agent (see equation 1).Finally, we take into account the Basal Ganglia functions of action selection by defining a set of stereotypical action sequences associated with different stages of the goal-oriented action process, such as foraging or homing as shown in [Friend and Kravitz, 2014]. Action sequences are triggered by contextual modules (see section 2.3.4) coherently with the definition of goals and sub-goals over the task time. In the current setup the action sequences associated with each goal were predefined.

Altogether, the adaptive layer is responsible for computing motivational drives and for action and reward values associated with individual items collected from the environment, providing an intermediate computational step between low level allostasis and high level decision-making. Moreover the adaptive layer is also responsible for implementing associative learning generating anticipatory actions based the acquired predictive value of environmental cues. In particular, the motor output of the adaptive layer, t_A, is integrated with the total motor response according to:

$$w_{left} = m_r + (t_{TN} + t_{SC} + t_A)$$

$$w_{right} = m_r - (t_{TN} + t_{SC} + t_A)$$

(15)

Contextual Layer

Episodic memory and path planning in the Hippocampus. Within the DAC contextual layer, the hippocampal module provides the computation for the robot spatial selectivity and navigational planning. A population of 720 head-direction cells was implemented as rate-based neurons in a ring attractor fashion [Eliasmith, 2005] and their activation function was defined by the robot's angular direction, provided by a digital compass device. Thus, every time the robot faced a specific orientation, cells tuned to that orientation became active. Six populations of 100 grid cells each were implemented as point neurons based on attractor dynamics and updated with motion signals provided by the soma, such as the robot's speed, and directionality, provided by head-direction cells activity [Guanella et al., 2007, Santos-Pata, Diogo Verschure, 2014, Yoon et al., 2013]. The scale of each population of grid cells was set to mimic the varying encodings found along the dorsal-ventral axis of the Medial Entorhinal Cortex (MEC) layer 2, such that grid cell firing fields distances range from 20 to 100 cm [Brun et al., 2008]. As in [Guanella et al., 2007], grid cells were initialized with random activity between 0 and 0.1 and are updated at each time step by a linear transfer function given by:

$$B_i(t+1) = A_i(t) + \sum_{j=1}^{N} A_j(t) w_{ij}$$

where N is the total number of neurons in the population of grid cells and w_{ij} represents the synaptic weight of each cell i to cell j, as described below. At the synaptic level, each neuron in a grid cell population is recurrently connected to every other neuron in the same population forming a toroidal topology with a synaptic strength defined by a Gaussian weight function as:

$$W_{ij} = I_{exp}(\frac{-|C_ij|^2}{\sigma^2}) - T$$

(17)

where c_{ij} is the distance from cell i to cell j in the network position. c_{ij} is transformed in order to maintain the toroidal topology. Synaptic strength is modulated by the intensity parameter I(=0.3). The Gaussian distribution is regulated by the parameter σ (=0.24), and excitatory/inhibitory zones are determined by the gaussian parameter T (=0.05). Modulation of synaptic weights was performed every time step determined by the robot's motion signals (all parameters are unit-less, see [Guanella et al., 2007] for more details).

In our hippocampal place cells implementation, a population of 1000 point neurons received excitatory input from randomly distributed synaptic connections arriving from all grid cell populations, as described in [de Almeida et al., 2009]. Thus, by convergence of spatially tuned grid cells, the resulting activity of each hippocampal cell becomes location specific. The excitatory input of each hippocampal cell is updated as:

$$E_{pc}(t) = \sum_{j=1}^{N_{grid}} W_{pc,j_{grid}} A_{j_{grid}}(t)$$

(18)

where the response, E_{pc}, of each hippocampal cell is the product of the rate of every MEC grid-cell, j_{grid}, with the specific synaptic weight, $W_{pc,j_{grid}}$. The firing rate of each hippocampal cell, A_{pc}, is the result of a competition process driven by MEC grid cells input within hippocampal cells governed by the $E\% - max$ winner-take-all process as in [de Almeida et al., 2009] and is defined as:

$$A_{pc}(t) = E_{pc}(t)H(E_{pc}(t) - (1-k)argmax_{pc}(t))$$

(19)

In our implementation, the value of $k(E\% - max)$ was defined as 0.1, such that place cells activate the firing of those cells whose excitation is within k maximal excitation of the entire population. H is the Heaviside function where $H(x)$ is equal to 1 if $x > 0$ and 0 otherwise.

During learning, synaptic weights between grid- and place-cells were susceptible to be reinforced or weakened through Hebbian mechanism. The probability of applying the Hebbian mechanism between grid- and place-cells was given by a decay function over time. In accordance with the experimental design, such decay function reach the value of zero at the end of early trials of the exrimental phase. Thus, at middle and late phases, synaptic weights between these two populations remained stable.

The hippocampal place cells become tuned to specific locations forming a short-term memory that allows the agent to generate its internal map of the environment (see results section). In order to generate short-term memory sequences and to be able to solve path planning, we set a connectivity matrix of $N2$ sweep cells (N = number of place cells) organized by previously active to currently active place cells [Johnson and Redish, 2007]. This connectivity matrix was initiated with synaptic weights set to 0 and every time two place-cells were active (moving from one place field to the next), the correspondent sweep cells was set to 1. As in [Milford and Schulz, 2014], we used Dijkstra’s algorithm to find the least number of sweep cells to be activated between the current and goal place-cell. After obtaining the shortest path, a set of vectors were computed to allow the agent to orient its heading direction. This was implemented with an orientation matrix of equal dimensions as the sweep cell connectivity matrix. Every time a sweep cell was updated, the correspondent angular orientation cell was set with the current heading direction of the robot. Thus, in order to reach a place-cell that has been associated with the target object, the sweep mechanism would retrieve a set of place cells as well as the angular directions to take in between place cells.

Decision-making and long-term memory in the medial Prefrontal Cortex Within the DAC-X architecture, decision-making occurs at the level of the contextual layer. This was modeled in reference to the dorsal medial Prefrontal Cortex, where behavioral plans are computed based on the integration of signals reflecting perceptual evidence, memory biasing and goal states (see Introduction). The decision-making model consists of three populations (pools) of excitatory neurons that have excitatory recurrent connections within themselves and that mutually inhibit other pools [Marcos, Encarni Verschure, 2013]. Each pool is sensitive to a specific drive-to-action signal arriving from the VTA module. The activity of each neural pool is updated according to:

$$\tau\frac{dU_i}{dt} = -U_i(t) + f(\lambda_i + \omega_+ U_i - \sum_m^N(\omega_{-m}U_m)) + \sigma\xi(t)$$

(20)

where U is the mean activity of the neural pool, λ represents the pool’s drive-to-action input signal, ω defines the strength of the connections of the network, ξ is a Gaussian noise (with mean 0 and variance 1) added to the network and modulated by σ (see parameter values below). The function $f(x)$ is a sigmoid function defined as:

$$f(x) = \frac{F_{max}}{1 + e^{\frac{-(x-\theta)}{k}}}$$

(21)

where F_{max} is the maximum value that the function can achieve, θ is the center of the sigmoid and k defines its slope. The decision process terminates when the mean activity of one of the neural pools reaches a decision bound. When a decision is made, the decision signal is sent to the long-term memory module. In our experiments, we used the parameters: τ=20ms, ω+ = 1, ω-m = 1 except when $m = i$ that ω-m = 0, σ=0.1 spikes s^{-1}, $F_{max} = 1.5$ spikes s^{-1}, θ = 4.44 spikes s^{-1} and $k = 0.4$ spikes s^{-1}.

In the presented framework, long-term memory was implemented based on the goal-place theory that associates goal-items with an internal representation of the explored environment [Duff and Verschure, 2010]. At the functional level, this module performs associations of objects with locations at which those objects are encountered. We have implemented the long-term memory system based on a location-cue buffer following the rule: every time the agent perceives an object that potentially satisfies internal needs, the most active hippocampal place cell is stored together with the encountered object at that same location. When a goal-specific decision signal arrives, the object-associated place-cell is retrieved and a hippocampal sweep sequence together with its encoded orientation is retrieved.

Altogether, the contextual layer in the present system is responsible for acquiring an internal representation of space, performing decision-making useful to define goals and store goal-location couplets defined in the hippocampal module. This layer also provides a form of goal-directed navigation based on internal representation of the environment used to navigate towards known locations. In particular, the motor output of the contextual layer , t_c, is integrated with the total motor output system and sent to the motors according to the following equation:

$$w_{left} = m_r + (t_{TN} + t_{SC} + t_A + t_C)$$
$$w_{right} = m_r - (t_{TN} + t_{SC} + t_A + t_C)$$

(22)

Finally in order to keep track of the motor performance of the agent we use a measure of motor cost. Coherently with [Fagg, 2002], computation of total motor cost, J_w, encoding the total effort performed by the agent at every trial and used for data analysis, is performed according to the following quadratic function:

$$J_w = \lambda \int_0^t (|w_{left}| + |w_{right}|)^2$$

(23)

where λ (=0.02) is a regularization parameter.

1.3 Results

In a first experiment we benchmark our system by looking at the evolution of learning across different phases of the task and across the multiple

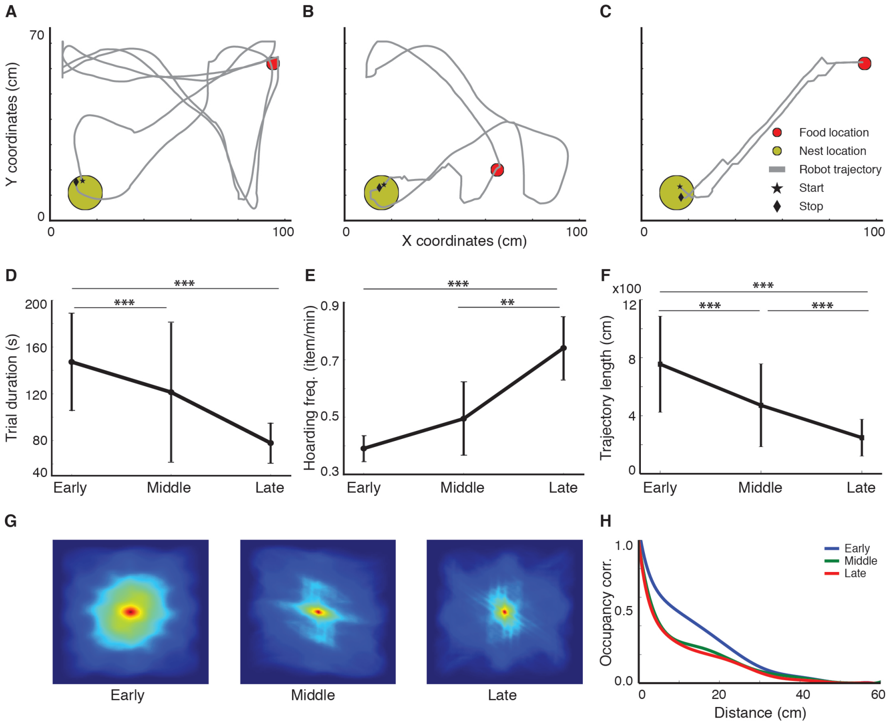


Figure 1.3: **Quantification of behavioral performance** for Early trials (trial 1-4; total trials = 36, 7 trials removed), Middle trials (trial 5-8; total trials = 36, 4 trials removed) and Late trials (trial 9-12; total trials = 36, 3 trials removed) for a total of 9 sessions. A, B, C. Example trajectory for one early, middle, late trial. Resource location refers to the encountered one during the trial. D. Mean trial duration and variance in seconds for different stages of learning. T-test significance for Early vs Late ($pvalue < 0.001$), Middle vs Late ($pvalue < 0.001$). E. Mean hoarding frequency and variance in items per minute for different stages of learning. T-test significance for Early vs Late (p value ¡ 0.001), Middle vs Late ($pvalue < 0.05$). F. Mean trial trajectory length and variance in centimeters for different stages of learning. T-test significance for Early vs Late ($pvalue < 0.001$), Middle vs Late ($pvalue < 0.001$), Early vs Middle ($pvalue < 0.001$). H. Occupancy correlation score in function of

distance to peak. As in G, higher correlation score along distance axis for early trials suggests homogenous occupancy within the arena, compared to more specific trajectories in middle and late trials.

layers of the architecture. We report our results by first analyzing the evolution of behavior, from early to late trials, for a total of 9 experimental sessions initialized at identical conditions. We discretize the behavioral performance across the whole duration of the task into three phases per each session: Early trials (session trial 1-4; total trials = 36 of which 7 trials were removed due to tracking system failure), Middle trials (session trial 5-8; total trials = 36 of which 4 trials were removed due to tracking system failure) and Late trials (session trial 9-12; total trials = 36 of which 3 trials were removed due to tracking system failure). We provide a quantification of behavior in terms of spatial (trajectory length, occupancy) and temporal (trial time, hoarding frequency) measures and analyze efficiency in terms minimization of motor cost and maximization of the value associated with the reward. Moreover we provide an insight in the internal dynamics of the principal modules of the architecture across different learning stages and for key behavioral events identified during the task. In a second experiment we provide a trial-by-trial quantification of the individual contribution of each layer of the control architecture to the behavior of the agent by progressively enabling each layer of the architecture starting from reactive (R, total sessions = 5; trials per session = 12), to reactive-adaptive (R+A, total sessions = 5; trials per session = 12) and, finally, to reactive-adaptive-contextual (R+A+C, total sessions = 5; trials per session = 12).

1.3.1 Quantification of behavior

At the beginning of every experimental session, the naÃ¯ve agent was required to explore the arena and seek for resources in order to satisfy a set of internal needs defined by the allostatic controller (hypothalamic nuclei) in the reactive layer of the architecture. The naÃ¯ve agent displays a reactive navigation strategy used to avoid walls and obstacles and to direct navigation towards visually located resources. Such navigation strategy, supporting exploratory behavior, is characterized by a stochastic trajectory pattern (fig. 1.3-A) that leads to a type of behavior characterized by high occupancy (fig. 1.3-G,H Early) and a longer trajectory length per trial (fig. 1.3-F Early) compared to later stages of learning.

This type of navigation is useful to locate resources in the environment due to the high coverage of the foraging arena, however it yields a relatively low item collection rate (fig. 1.3-E Early) compared to more advanced stages of learning, and a high mean trial time (fig. 1.3-D Early). To measure navigation redundancy we have obtained the occupancy maps from trajectories at each navigation phase. Redundancy is defined by the coverage of the arena by the trajectory, i.e. revisiting the same location multiple times within the same trial, and is independent of trajectory length. Occupancy maps were cross-correlated in order to measure the spatial tenancy (fig. 1.3-G Early). Linearization of the mean occupancy correlation in function of peak-distance (fig. 1.3-H Early) reveals that early trials require high level of navigational exploration in order to complete a trial. Overall, this strategy is necessary to explore the environment, locate resources and to build an internal representation of space; however behavioral performance in terms of goal-achievement remains low.

During an intermediate stage of learning, the initial reactive navigation strategy is complemented with adaptive responses useful to perform more efficient obstacle avoidance and to make use of local visual landmarks (green patches on the floor) that facilitate anticipatory movements towards a resource location (see section 3.3.1). Moreover the initial exploratory pattern fostered the consolidation of an internal representation of space that supports goal directed navigation towards known locations (i.e. home location). During this phase of learning we report a mixed trajectory pattern (fig. 1.3-B) as the result of reactive exploration, cue-based navigation and, in some cases, goal-oriented navigation towards resource or home location. Such strategy yields a reduced occupancy of the arena (fig. 1.3-G,H Middle), suggesting a progressive shift from purely exploration to exploration-exploitation type of behaviors. This is confirmed by the reduction of the mean trajectory length in this phase (fig. 1.3-F Middle). The improved navigation strategy in this phase influences hoarding behavior, producing a higher collection rate (fig. 1.3-E Middle) and a lower mean time per trial (fig. 1.3-D Middle). Occupancy correlation was minimized during Middle trials, which can be explained by the fact that during early trials place-cells are susceptible to learning and thus navigation is not supported by a stable spatial representation (fig. 1.3-G,H Middle).

Finally, during late stages of learning, the navigation strategy adopted by the agent was enhanced by a complete contextual knowledge of the environment. Such goal-oriented strategy is mostly driven by a combination of a robust representation of the environment (see section 3.3.2) and the availability of goal locations stored in long-term memory, used to perform shortest path navigation towards resources. This phase of learning was characterized by a goal-oriented behavior that yields mostly linear trajectory patterns (fig. 1.3-C) leading the agent to a resource location and back to the home location. The occupancy measure maintained a low correlation score (fig. 1.3-G,H Late) and a shortest path length (fig. 1.3-F Late), suggesting a navigation strategy driven by exploitation of acquired information with a direct effect on hoarding performance, that is expressed in the high collection rate (fig. 1.3-E Late) and a reduced mean time per trial (fig. 1.3-D Late).

1.3.2 Quantification of performance

The above measures are useful to describe the behavioral performance of the DAC-X in the hoarding task, however they do not allow us to fully analyze the intrinsic performance of the system. For example, trajectory length and occupancy are certainly related with spatial measures but they don’t explain the motor effort expended by the agent, which might cover similar trajectories performing motor commands of different magnitudes depending on the stage of motor learning (i.e. reactive vs adaptive obstacle avoidance). Similarly, mean trial time and collection rate are related to the amount of reward obtained by the agent, but they don’t unambiguously account for its goal achievement due to the temporal parabolic discounting applied to the reward value (see Methods).

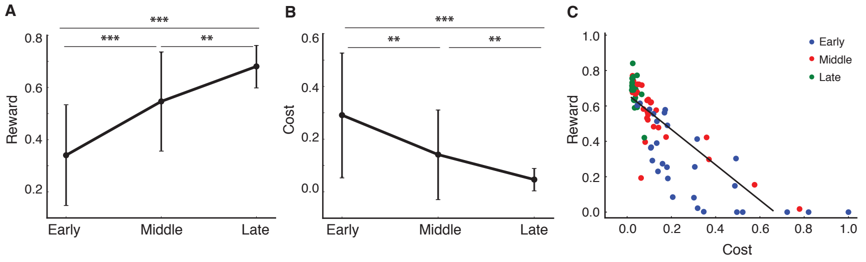


Figure 1.4: **Quantification of efficiency** expressed in terms of motor cost and reward per trial for Early trials (trial 1-4; total trials = 36, 7 trials removed), Middle trials (trial 5-8; total trials = 36, 4 trials removed) and Late trials (trial 9-12; total trials = 36, 3 trials removed) for a total of 9 sessions. **A.** Mean reward value and variance per trial (normalized by the maximum reward value of individual item) for different stages of learning. T-test significance for Early vs Late (p value ¡ 0.001), Middle vs Late $(pvalue < 0.01)$, Middle vs Early $(pvalue < 0.001)$. **B.** Mean cumulative motor cost and variance per trial (normalized by the maximum of every session) for different stages of learning. T-test significance for Early vs Late $(pvalue < 0.001)$, Middle vs Late $(pvalue < 0.01)$, Middle vs Early $(pvalue < 0.01)$. **C.** Cumulative motor cost over reward value per trial in Early (blue), Middle (red) and Late (green) stages of learning. Pearson’s correlation coefficient: $r : -0.87, p - value < 0.01$.

We further quantified the performance of the agent by looking at motor cost, reward value and their relationship during the evolution of learning as a more reliable measure of hoarding efficiency. Along the progression of learning during the performance of the task we notice a continuous decrease of motor cost (fig. 1.4-B). This improvement is the result of learning within the adaptive layer, which provides more efficient obstacle avoidance (see section 3.3.1) together with an increased probability to encounter resources. In addition, the contextual layer (see section 3.3.2 and 3.3.3) provides an acquired strategy for efficient goal-oriented navigation. An opposed trend can be found in the development of the reward value associated with each item, which increases with the advancement of learning (fig. 1.4-A). Together, these opposite trends hold a linear relationship suggesting a tendency towards optimal hoarding (fig. 1.4-C).

1.3.3 Single module dynamics and interaction

In order to outline the intrinsic processes that underlie the behavior of the agent, we propose a more detailed analysis on single module dynamics. First, we analyze the acquisition of adaptive motor responses in the adaptive layer and spatial memory in the contextual layer. We also provide an insight in the decision-making process performed by the prefrontal module. Finally we provide an overview on the interaction of the multiple processes that generate behavior for key events during the performance of the task.

Cerebellar modules

We implemented a set of microcircuits responsible for the acquisition of avoidance responses that learn from the output of the reactive layer and are used to increase the ability of the agent to prevent collisions with walls and obstacles through acquired avoidance. This form of motor learning allows the agent to perform turning actions that anticipate the output of the reactive controller, resulting in safer and smoother turns. Such improvement can be noted in (fig. 1.5-A) where an initial reactive response (Naive), triggered by the high proximity to a wall, is replaced, after learning, by an anticipated adaptive response characterized by lower amplitude (Trained). For this reason, adaptive obstacle avoidance allows for a less costly and more effective motor strategy and could play an important role in increasing the efficiency of the agent in the execution of the task.

A second set of microcircuits was dedicated to exploit the acquired predictive value of landmarks in the environment in order to perform a simple form of cue-based navigation towards resource sites. In particular, the perception of a visual cue is progressively associated with an efferent copy of an attentional reactive response so that, after a number of repetitions (mean=5 for all the sessions) the cue is sufficient to trigger a turning

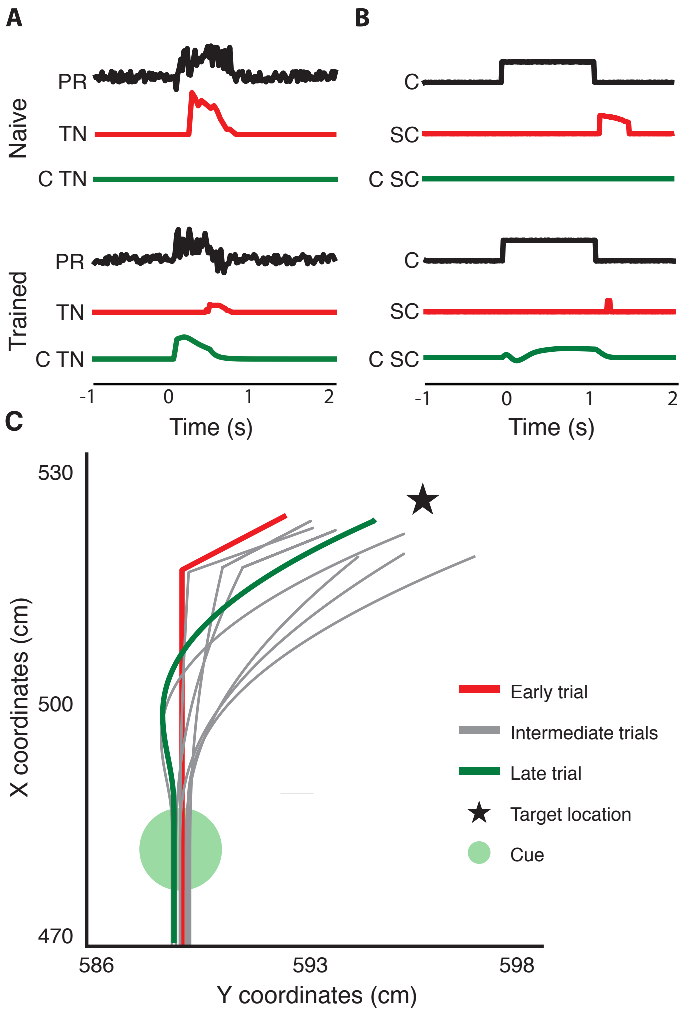


Figure 1.5: **Adaptive responses in the cerebellar modules. A.** Example of adaptive avoidance response learning. Naive trial (occurrence n=1): the activity of the proximity signal (PR) triggers a reactive response in the feedback reactive controller (TN) peaking at amplitude = 1.2. The adaptive response (C TN) in this trial is absent. Trained trial (occurrence n=5): the activity of the proximity signal (PR) triggers an adaptive response (C TN) that precedes the reactive response (TN) and allows longer range

turning. Note that C TN response peaks at amplitude = 0.7 allowing more efficient obstacle avoidance compared to NaÃ¯ve trial. **B.** Example of cue response association underling landmark navigation. NaÃ¯ve trial (occurrence n=1): the activity of the camera indicating a patch on the floor (C) is followed in time by a reactive response (SC) directing the robot towards a perceived resource. The adaptive response (C SC) in this trial is absent. Trained trial (occurrence n=5): the activity of the camera indicating the patch on the floor (C) now triggers an adaptive response (C SC) as the result of an acquired cue-response association, that precedes in time the reactive response (SC) and allows anticipatory turning. **C.** Behavioral outcome effect of adaptive turning: trajectories for reactive turning (red) and adaptive turning (green) strategies. Gray trajectories represent intermediate occurrences. Green circle represents the visual cue. Star marker represents the resource location.

action towards a previously encountered location. Such dynamic can be appreciated in (fig. 1.5-B) where a comparison between pre-learning and after-learning phases shows how the reactive response strongly present in early trials (Naive) is replaced, in late trials (Trained), by an acquired adaptive response triggered by the visual signal. Such adaptive motor response ultimately allows the agent to navigate by making use of local cues and to increase the probability to encounter resources in the arena (fig. 1.5-C). Finally, this form of learning can contribute to the overall performance by making the search process faster and more effective and indirectly affecting the amount of reward value obtained at every trial.

Hippocampal modules

Within the hippocampal formation, three modules allowed the agent to build an internal representation of space. Grid cells served as the core metric system and each cell was characterized by its grid-scale and orientation (fig. 1.6-C). Grid cells allowed place cells to receive multiple randomly distributed inputs and served as their primary activation function. The place cells winner-take-all mechanism modulated their behavior, leading to position specific activity. During early trials, place cells formed broader and unspecific rate maps when compared to late trials (fig. 1.6-A). In order to quantify place fields, a 15-pixels radius Gaussian kernel was used to smooth the rate maps and clusters of neighboring pixels above 10% of maximum activity were considered a place field. Mean and variance of place fields number per cell was higher at early trials compared to late trials, when most cells had one place field and few had zero or two place fields (fig. 1.6-B). Thus as the robot explores, place cells become more tuned to unique locations.

Place cells activity modulated sweep cells interconnectivity strengthen with spatially proximal cells (see Methods section). During goal-directed behavior, the sweep mechanism activated a sequence of cells that were spatially contiguous through their associated place cells (see examples of sweeps activity and required place cells in fig. 1.6-C). Synaptic updates in sweep-to-sweep connectivity tend to stabilize as the agent explores the environment (fig. 1.6-E) and place cells became position specific (fig. 1.6-B).

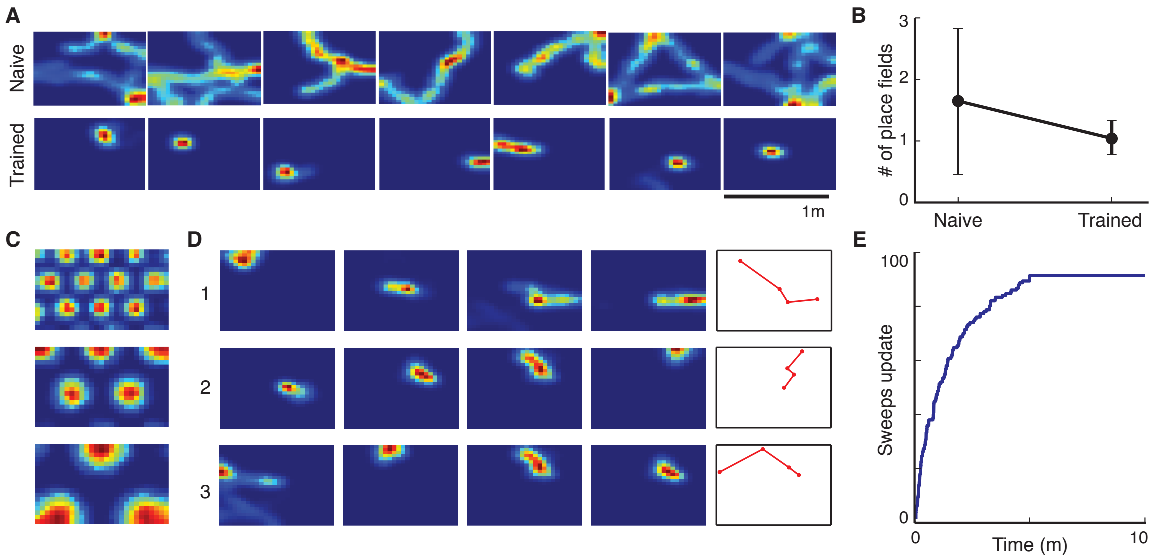


Figure 1.6: **Spatial memory in the hippocampal modules. A** Examples of rate maps from NaÃ¯ve and Trained place cells activity. NaÃ¯ve place cells show diffused and broader activity along the explored arena, while trained cells are mostly location specific.**B** Quantification of number of place fields per cells indicate that after training, place cells became tuned to unique locations (mean:1.06, std:1,16) compared to naÃ¯ve cells (mean:1.64, std:0.23) . **C** Grid-cells rate maps from cells with different grid-scale from three dorsal-ventral axis levels. **D** Examples rate maps from recruited place cells during three goal-direct behavior. Left column shows rate maps of active place cell when sweep event was triggered. Subsequent rate maps illustrate recruited cells to reach goal place cell. Vectored trajectories are illustrated on last

column, where dots represent cells rate map center of mass. **E** Modification of sweep cells connectivity during session. Rapid modifications at the beginning of each session tend to stabilize as the agent forms its internal map.

Prefrontal modules

In DAC-X an important role is played by the decision-making and long term memory functions performed by the Prefrontal modules. Decision-making dynamics are accounted for by the interaction among different neural pools responsible for resolving the competition between different motivational drives conveyed by the VTA module at the beginning of every trial. This mechanism allows the agent to establish a goal, namely a target resource to be fetched, at the beginning of every trial. An example of this decision process can be seen in (fig. 1.7-A), where the rising activity of every pool accounts for a competing goal relevant for the task. Competition is resolved when one of the pools reaches enough activity to inhibit the competitors and cross a decision threshold (here set to 1.0), above which the goal is selected. In the present decision-making model competition was influenced by both magnitudes of the input, such that a greater activity encoded by the VTA module would increase the probability of that pool winning. A second factor influencing the decision-making process can be found in the intrinsic dynamics of each pool, where spontaneous noise adds variability to the competition so that for equal input magnitude the network might fall into different attractor states (fig. 1.7-B).

Another computational feature derived from the prefrontal module is the long-term memory of goal-location couplets. In the current setup, memory is defined as a structure associating the hippocampal activity of specific place cells with an encountered resource location and it contributes to the spatial knowledge exploited by the agent in the hoarding process. Importantly, this feature is used by the hippocampal sweep mechanism (see section 3.3.2) in order to define a path toward a goal location. In (fig. 1.7-C) we show the increasing amount of goal-place couplets stored in memory as an important underling factor to the performance of the robot. These couplets, due to the sweep mechanisms thus also implicitly encode the heading vector that must be followed to reach the goal location.

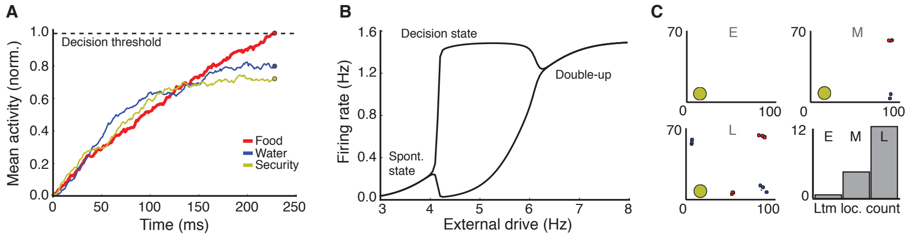


Figure 1.7: **Decision-making and long-term memory in the prefrontal modules. A.** Example of competition among multiple choices triggered at the beginning of every trial by the prefrontal module expressed in normalized pool activity. Each line represents the encoding of a choice among food seeking (red), water seeking (blue) and security seeking (yellow). In the present example food-seeking behavior is selected when the total pool activity reaches the value of 1.0 (decision threshold). **B.** Bifurcation diagram of the PFC network for the selected parameters. The figure shows the different states in which the model can operate depending on the external drive provided to the three selective pools of the network. When the network operates in the Decision State only one neural pool fires at high rate whereas the other two fire at a low level rate. In the other two possible states the three pools remain at the same level of activation that depends on the specific range of the external drive: low or high level for spontaneous and double-up, respectively. **C.** Long term-memory representation (converted to x and y coordinates expressed in centimeters) of encountered food (red) and water (blue) resource locations across different stages of learning, from Early (top-left panel), Middle (top-right panel) and Late (bottom-left panel) for a total of 9 sessions. Yellow circle represents the home location. A total count of LTM locations is presented in the bottom-right panel for the same stages of learning.

Architecture dynamics for key events

To provide an overview of the dynamics of the interaction of multiple modules within the architecture we selected key events during the execution of the task highlighting the contribution of each process with respect to the

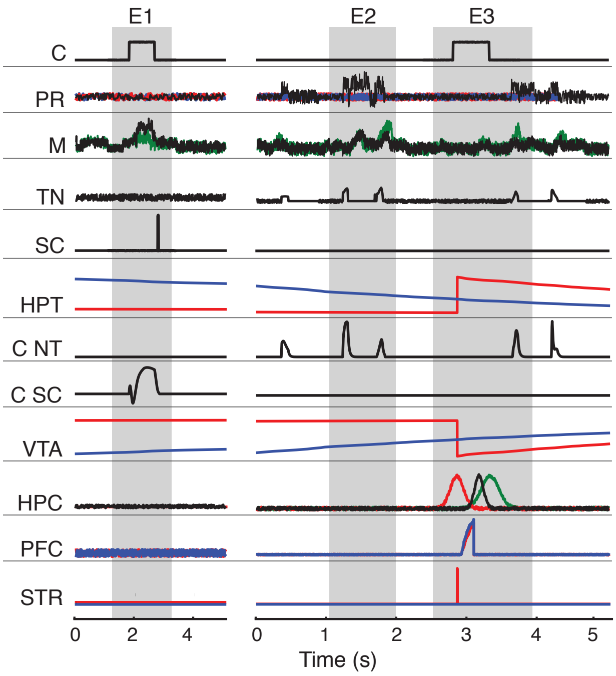


Figure 1.8: **Architecture activity for significant events. E1.** Cue-response association underlying cue-based navigation strategy. Note the concomitance of increased activity in perceptual modules (camera (C)), motor modules (M; motor left, black; motor right, green), reactive modules (SC) and cerebellar module (C SC). **E2.** Obstacle avoidance. Note the concomitance of increased activity in perceptual modules (proximity sensors (PR)), motor modules (M; motor left, black; motor right, green), reactive modules (TN) and cerebellar module (C TN). **E3.** Reward delivery and decision-making. Note the concomitance of activity of the perceptual modules (camera (C)) perceiving home location (yellow patch), together with the activity in selected cells of the hippocampal module (HPC) encoding for the same location. At the delivery of the food reward (STR, red) hypothalamic values associated with food (HPT, red) are restored and motivation in VTA modules decreased (VTA, red), triggering a decision in pre-frontal modules for water-seeking goal. (PFC, blue).

others. In particular, we select an example of adaptive cue-based navigation and adaptive avoidance responses, an example of reward achievement and decision-making and finally, an example of arrival at resource location. Adaptive cue-based navigation (fig. 1.8-E1) is characterized by the perception of a visual cue (C) during the exploration process which triggers an adaptive motor response (C SC) that conveyed to the motors (M) allows the agent to turn toward a resource location. Such a response partially overwrites the reactive attentional response (SC) by acting in anticipation. Similarly avoidance responses (C TN - fig. 1.8-E2) triggered by proximity sensor signal (PR) allow the robot to efficiently prevent collisions with obstacles. When the agent reaches the home location (fig. 1.8-E3), visually identified by a landmark (C) as well as by internal spatial representation (HPC), a reward value associated with the hoarded item is delivered (STR). Delivery of reward affected the allostatic value for the addressed internal state encoded in (HPC), increasing its values and consequently decreasing the motivation to pursue that type of resource encoded in (VTA). Activity in VTA modules directly affect the decision-making process (PFC) performed at the beginning of the trial, biasing the decision towards a most urgent need.

1.3.4 Contribution of individual layers to behavior

We conclude the report of our results addressing the question of what each layer of the DAC-X architecture contributes to the overall performance of the agent in the execution of the task. To do this we ran a set of 5 sessions (12 trials per session) for three different conditions, where we progressively enable for each condition a layer of the architecture, starting from reactive only (R), followed by reactive and adaptive (R+A) and finally reactive, adaptive and contextual (R+A+C). We report the overall evolution of the hoarding efficiency in a trial-by-trial fashion, quantified in terms of reward-cost ratio.

The form of control provided by the reactive layer is entirely based on reflex-like motor responses and primitive attentional mechanisms. Nevertheless, the agent is capable to execute the task and hoard resources albeit with a high level of variability in performance. Motor cost (fig. 1.9-A,R and fig. 1.9-D blue for a single trial example) for a given trial varies in a wide range of values, coherently with the amount of reward obtained per item (fig. 1.9-B,R). Such variability can be attributed to the non-goal-oriented navigation strategy that essentially provides variable chances to encounter reward depending on the variable feedback provided by the sensors. Not surprisingly, hoarding efficiency is not subject to appreciable changes along the whole session due to the absence of learning in the reactive layer (fig. 1.9-C,R). Such strategy yields a relatively flat learning curve (mean value = 2.8) that slightly decreases as available resources are hoarded and the chances to encounter new ones decrease.

Enabling the adaptive layer of behavior introduces the first element of associative learning in the control architecture. Learning in this layer can be found in the adaptive properties of the cerebellar modules useful to perform less costly collision avoidance combined with simple cue-based navigation. An effect of this kind of learning is a decrease of mean motor cost (fig. 1.9-A,R+A and fig. 1.9-D red for a single trial example) and its associated variance as well as an increase of mean reward value (fig. 1.9-B,R+A). For reward value however variability is equal to condition R, reflecting a hoarding strategy that is still mostly driven by sensory contingencies. Overall, the introduction of the adaptive layer leads to an increase in performance over trials (fig. 1.9-C,R+A). Initially condition R+A does not differ from R, however an increase in efficiency can be appreciated from the third trial on, where use of local cues for navigation towards resources increases the probability to collect a higher reward. Performance, which reaches a peak value of 6, is by necessity variable and the decrease during the course of the session is due to the progressive depletion of resources, i.e. lacking the capability to plan: in condition R+A the agent can only optimize performance in a restricted spatiotemporal window making it less likely for it to collect more difficult targets.

The reported performance of the adaptive layer partially depends on the parameters configuration chosen for the task. In particular we set a learning rate (β = 10 see Methods 2.2.3) which allowed the agent to acquire a mature adaptive response within a temporal window of 2-3 trials without inducing instability of the cerebellar output. Another key parameter of the cerebellar module can be found in the delay, which informally governs the anticipatory time span of the adaptive response over the reactive response encoded in the teaching signal. In the current setup the delay, δ, was fine-tuned to 800 ms (see Methods 2.2.3) in order to guarantee an appropriate anticipatory range given the inertia of the servo motors. A shorter delay would have been incompatible with the dynamics of the robot by producing no anticipation, while a longer value would have produced greater anticipation and possibly causing over-shooting actions, thereby affecting performance. Finally the K_noi parameter, representing the gain of the nucleo-olivary inhibitory synapses from the deep cerebellar nucleus to the Inferior Olive, controls the magnitude of the teaching signal conveyed to the cerebellar cortex, informally balancing the amount of adaptive and reactive signals forming the final motor response. In this particular setup we set this parameter to 0 because the generation of the error can be fully accounted by behavioral feedback. This means that the error can be fully canceled by behavioral response without the need of internal error generation. Setting the K_noi to a higher value might improve robustness and responsiveness of the system during extinction paradigms (see [Herreros et al., 2013b]) which are outside the scope of the present study.

The contextual layer of control introduces spatial learning, decision-making and long term memory features to the behavior of the agent. This layer greatly increases the performance of the agent, which is now able to represent and efficiently navigate the arena as well as remember relevant locations. We observe a highly significant decrease in motor cost (fig. 1.9-A,R+A+C and fig. 1.9-D green for a single trial example), a mean value of obtained reward comparable to the R+A condition with a reduction, however, of its variance (fig. 1.9-B,R+A+C). This suggests that the main contribution of the contextual layer, i.e. goal-oriented planning, is in the ability to save motor energy providing shorter paths towards desired location. This strategy indirectly optimizes reward, likely given by the reduced variability in the time spent to hoard a resource. A positive learning curve shows an increase of overall efficiency during the course of the sessions reaching a peak value of 13 (fig. 1.9-C,R+A+C). We expect the difference in obtained reward to become more prominent in more complex tasks.

The observed performance with the addition of the contextual layer is tightly related to the initialization parameters of the hippocampal complex and pre-frontal areas. The details of the integrated hippocampal model were sufficient to match anatomical, physiological and functional properties of its biological counterpart. Despite its core computations, the model dynamics were tuned to fit the biological constraints found in both grid- and place-cells of the rodent Hippocampus. Specifically, the grid-cell gain parameters were set to threshold grid-scale at every module such that firing fields were within a range of 20 cm to 100 cm distanced of each other, as found in the rodent MEC dorsal-ventral axis [Brun et al., 2008]. The synaptic projections from grid- onto place-cells were randomly defined with weights from 0.0 (no connectivity) to 1.0 (full connectivity). Further, Hebbian dynamics shaped these weights during exploration. The rule for Hebbian activation was probabilistically defined by a decay function over time during early exploration phases, and only applied when more than 50% of place-cells showed activity grater than to 75% of hippocampal maximum activity. This mechanism allowed the tuning of place-cells location sensitivity and stability over sessions. The E%-max winner-take-all mechanism was also used to realize competitive dynamics within the place-cells population. As in [de Almeida et al., 2009] we have set a 10% threshold of the maximal population activity. However, the consequences of parameter modulation was not considered in the present study. Altering the grid-scale, for instance, would have an enormous impact on goal-oriented behavior performance, given that it would modulate the resolution at which grid-cells encode space. The grid-to-place-cell synaptic learning mechanism would also impact the rate at which place-cells become location specific, affecting the formation of the spatial representation. Lastly, the suprathreshold rule for place-cell competition was at a fixed spatial range. However, altering the amount of cells competing for a specific location would also affect the dynamics and stabilization of the hippocampal spatial representation.

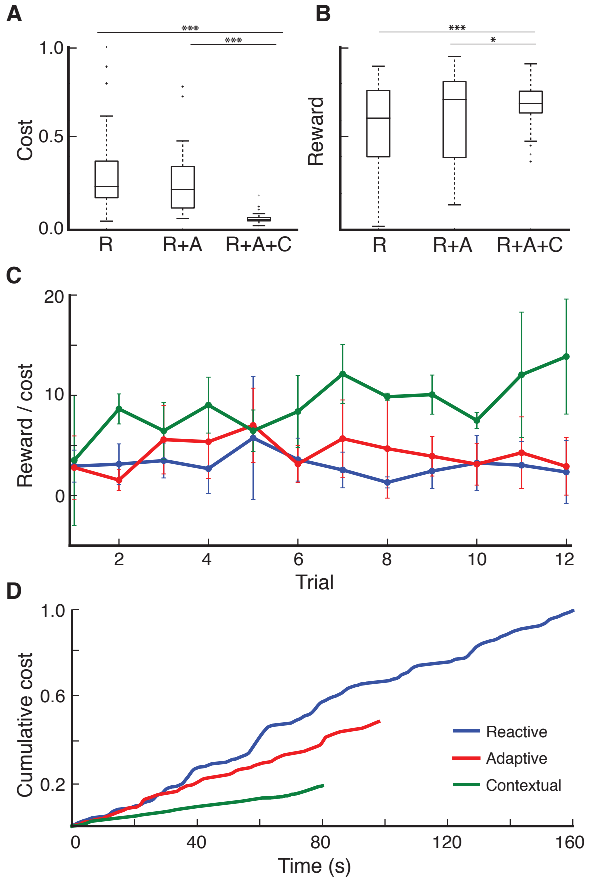


Figure 1.9: **Contribution of individual layers of behavior**for three different conditions with only Reactive layer enabled (R), Reactive and Adaptive layers enabled (R+A) and Reactive, Adaptive and Contextual (R+A+C) for a total of 5 sessions (total trials=60) per condition. A. Distribution of total cumulative motor cost per trial (normalized by the maximum of every condition) for different conditions. T-test significance for R vs R+A+C (p value ¡ 0.001), R+A vs R+A+C (p value ¡ 0.001). B. Distribution of reward value obtained per trial (normalized by the maximum reward value of individual item) for different conditions. T-test significance for R vs R+A+C (p value ¡ 0.001), R+A vs R+A+C (p value ¡ 0.05). C. Evolution

of mean efficiency expressed in normalized reward-normalized cost ratio for three different conditions (R, blue; R+A, red; R+A+C, green) over the course of 12 trials of a session. Each data point represents the mean and variance in performance of the same trial number for different sessions for a total of 5 sessions. D. Example of cumulative cost for a single selected trial for each condition (color code as in C).

Together with the hippocampal formation, the Prefrontal decision making module within the contextual layer realizes a WTA dynamics for goal definition in the presence of multiple competing motivational signals. The strength of the competition is controlled by the external drive provided to the goal pools which is given in a range within $\tilde{4}$ and $\tilde{6}$ Hz to guarantee the termination of the competition when one pool reaches threshold (see fig. 1.7-B). Here the decision threshold was set to an arbitrary value of 1 for the normalized activity of the pools. These parameters, together with the modulation of the noise distribution, σ, affect the decision dynamics in two aspects: decision time and decision accuracy, or how a prioritized goal matches the relevant need. The first component affects performance by modulating reward discounting, while the second affects reward achievement. Here we chose values for these parameters that minimize errors in decision accuracy and guarantee an overall acceptable speed-accuracy trade-off. Modulation of the decision threshold (i.e by an urgency signal) and the introduction of decision biases due to learning would complexify the dynamics. This type of analysis, however, is outside the scope of the present paper but will be certainly be addressed in future studies.

In summary, we conclude that a progressive increase of efficiency can be detected for every condition, where learning plays an important role at every layer. However, it is the contextual layer that contributes the most to performance for this task by complementing reactive and adaptive navigation strategies with a shortest-path type of strategy and providing the ability to reuse previously acquired knowledge to optimize the cost/reward ratio.

1.4 Discussion

We addressed the question whether foraging behavior can be understood in terms of the neuronal substrate of classical and operant conditioning. Following the Distributed Adaptive Control theory of mind and brain we have decomposed foraging in terms of the H4W problem and we have mapped each of its objectives to specific neuronal structures. The brain functions addressing challenges such as satisfaction of internal needs, spatial navigation and the use of environmental cues for gathering and storing resources have been described in multiple studies [Gould et al., 2010, Keen-Rhinehart et al., 2010, Smulders et al., 2010]. However, the organizational principles structuring these functions in a coherent control architecture generating complex behavior such as foraging are still unclear. It is especially here that our study makes its contribution. We have implemented and tested the previously identified neural correlates of the DAC architecture within a single computational framework called DAC-X. At the implementation level, we integrated several computational models of different brain regions controlling the behavior of a real-world mobile robot performing a foraging task, where the properties of the environment had to be learnt in order to succeed. At the beginning of each session, a naÃ¯ve robot driven by its internal needs was placed into an unknown arena. In order to optimize the task, each model had to improve its learning mechanism for a specific objective-function, such as: 1) learning multiple sensory-sensory and sensory-motor associations in order to anticipate adaptive responses; 2) generating an internal map of the environment to successfully navigate; 3) storing and maintaining goal-place couplets in LTM; 4) generating navigational plans to achieve a goal location. Working together, the learning mechanisms of each model led to the maximization of reward and minimization of motor cost during each session, suggesting that both serial and parallel processing mechanisms must occur in order to optimize behavior. Moreover, our results suggest that the fundamental H4W problem can be solved by the principles underlying classical and operant conditioning.

DAC follows a method of convergent validation, where a model is asked to satisfy behavioral, anatomical and physiological constraints [Verschure, 2012]. A synthetic approach towards understanding mind, brain and behavior has a long history going back at least to Tolman and Gray Walters (for review see [Cordeschi, 2002]). Since the late 80ies Gerald Edelman has advanced the, so called, Darwin series of real-world brain based devices, arguing that this synthetic method is a key step in advancing our understanding of how behavior is generated by the system level interaction of multiple artificial neuromimetic controllers [Edelman, 2007]. DAC-X is consistent with this proposal but it deviates in the specific theory it advances to explain complex behavior such as foraging and the way in which it decomposes the behavior into the underlying H4W objectives. In order to demonstrate the behavioral outcome of multiple neural structures working in synergy, we implemented a biologically plausible cognitive architecture, named DAC-X, which controls a robotic agent engaging in a hoarding task. We have shown that the agent can efficiently perform in this setup by learning from the interaction with its surroundings and progressively increase its performance based on the learning and problem solving principles found in classical and operant conditioning. According to the definition of classical conditioning, an innate response, i.e. a reflex, becomes expressed in a response to a previously neutral stimulus due to their contiguity. Here we have shown how the interaction between the reactive and adaptive layers of DAC-X leads to behavioral outcomes consistent with the paradigm of classical conditioning. For instance, collision avoidance in this setup can be understood in terms of avoidance learning, where an animal issues an anticipatory response to avoid a predictable noxious stimulus [Gormezano et al.,]. Predictive attentional responses, on the other hand, can be compared to preparatory actions such as those found in conditioning paradigms where anticipatory feeding responses (i.e. salivation or mastication) are triggered by the presence of predictive sensory stimuli [Woods, 1991, Woods and Ramsay, 2000]. Overall the adaptation provided by these two layers of DAC-X, might account for some of the cognitive aspects found in animal foraging such as the use of cues for navigation and, more generally, ’foraging innovation’ [Dugatkin, 2008].

The use of classical conditioning paradigms in robots by embedding biologically plausible models of the Cerebellum within their control architectures is not new [Van Der Smagt, 1998]. For example, it has been shown how this form of learning can lead to enhanced motor control in both simulated [HofstoÈtter et al., 2002, Herreros and Verschure, 2013a] and real world agents [McKinstry et al., 2006, Herreros et al., 2013a] engaged in an adaptive maze navigation task. This form of adaptive control however has been often presented in isolation without addressing the interaction with other modules mimicking complementary brain functions (see [Inderbitzin et al., 2010] for an exception). Here for the first time, we integrated cerebellar learning dynamics in the broader context of a realistic real-world task and showed how the versatility of cerebellar learning can be employed in multiple aspects of sensory-motor learning partially contributing to the global performance of the agent.

Additionally, we integrated classical conditioning with operant conditioning. In operant conditioning the animal learns to produce a given action because of its consequences, i.e. producing a reward [Mackintosh, 1983]. Operant conditioning requires a more sophisticated form of learning where reward seeking behaviors driven by intrinsic motivation lead to context dependent decisions, such as in setups where an animal is required to navigate to a specific location and perform an action to gain a reward [Jurado-Parras et al., 2013]. Within a foraging task, this can be observed, for instance, when an animal performs hoarding behaviors requiring to learn specific contextual associations between a location and a reward (i.e. resource location) [Luo et al., 2011]. In the presented task, operant conditioning involves multiple cognitive and sensory-motor processes and the interaction of multiple sub-systems. The Hippocampus is mainly involved in contextual learning and spatial navigation [Burgess et al., 2002] and receives spatial motion related signals from the medial Entorhinal cortex [Knierim et al., 2014] as well as higher-level inputs from Prefrontal areas [Ito et al., 2015]. Indeed, changes in synaptic strength in hippocampal sub-regions during operant conditioning in foraging-like tasks have been reported [Corbit and Balleine, 2000, Jurado-Parras et al., 2013]. Specifically, high amplitude changes in excitatory postsynaptic potentials in the CA3 and CA1 regions are observed in appetitive behaviors.

Recently a number of robotic implementations capitalized on the unique computational role attributed to the Hippocampus in order to create artificial controllers, inspired by hippocampal anatomy and physiology, able to generate grid cell-like and place cell-like spatial representations. Krichmar et al. [Krichmar et al., 2005] for example proposed a brain-based device engaged in a spatial memory task, comparable to the Morris water maze, where a neural model of the thalamo-hippocampal circuit, in conjunction with simulated cortical sensory areas, was used to learn to navigate to a target location. Also [Waniek et al., 2013] have presented an integrated system of grid- and place-cells for robot navigation and localization in a massively-parallel computing architecture. Similarly, [Erdem and Hasselmo, 2014] have presented a hierarchical goal directed navigation model where multiple hippocampal-like cells were integrated to form spatial representation and probe linear look-ahead trajectories in a simulated environment.

The problem of Self Localization and Mapping (SLAM) has long been a priority in robot navigation. Such an intrinsically probabilistic task aims for a mobile robot to be able to represent spatial information and access relationships between cues found in the environment [Smith et al., 1990]. Ideally, the robot starts at an unknown location within an unknown environment and it is still able to estimate its absolute position by building a spatial representation of its surroundings. This problem is often solved by the engineering solution of sensory integration and filtering [Dissanayake et al., 2001] leading to robust levels of spatial accuracy. Despite the advantages of such an approach in spatial localization precision, the constraints faced by biological systems to solve the SLAM problem are often neglected. Indeed, the animal sensorial apparatus is limited in its capabilities when compared to millimeter-wave radars, for instance. In our approach, however, sensory signals used for the formation of the agent's spatial representation were approximated by the type of signals found in biology, such as head-orientation, acceleration or visual inputs. Regardless of these sensorial limitations, our hippocampal model was still able to form a map of the environments sufficient to support goal oriented navigation. Additionally, the increasing resolution and stability of the hippocampal place fields over the session led to an improvement of further higher-level processes such as long-term memory storage and spatial planning. Hence, improvements in performance result from the co-development of sub-systems in the task rather than by the competence of single specialized systems.

Our approach is comparable to previous attempts to model hippocampus based navigation, to the extent that the formation of place-fields encoding unique locations within an open arena were used as spatial anchor points supporting efficient map based navigation. The present implementation however differs from [Krichmar et al., 2005] in that our agent’s behavior is modulated by the integration of multiple sub-cortical and cortical regions synergistically working to satisfy emulated internal needs. Moreover, similarly to [Erdem and Hasselmo, 2014], we propose a navigation strategy where spatial encoding is an emergent property of the hippocampal sub-system acquired by experience. However, in order to form sensory-place memories, we integrate sensory inputs within a neural substrate coherent with the hippocampal anatomy and physiology. The emergent map is then used to perform shortest path approximation, using a similar technique as found in [Milford and Schulz, 2014]. Finally, differently from the previously proposed models, the notion of goal-state was included within the hippocampal computation. This feature, coherent with recent physiological findings suggesting Prefrontal projections to CA1 via nucleus reuniens [Ito et al., 2015], allows the Hippocampus structure to be informed on the current goal state of the system and supports goal directed navigation establishing an objective resource location to be reached.

The putative role of the Prefrontal Cortex in decision-making has been captured by competition dynamics among neuronal pools, orchestrating internal conflict resolution (i.e. what internal drive/goal should be prioritized) at the highest layer of the architecture. Such type of computation, coherent with neural dynamics of perceptual decision-making might partially explain the initiation of goal-oriented behaviors found in operant conditioning. Decision-making processes in the Prefrontal Cortex indeed defined the goal of the agent at the beginning of every trial and they were further used to instantiate a coherent action plan by accessing long-term memory structures. Here, behavioral planning was defined as the navigational strategy recruiting sequences of hippocampal place fields describing the path towards the goal location and further actuated by action selection processes. This is consistent with the type of goal directed navigation found in animal foraging [Gould et al., 2010]. Within the decision-making literature, several models have been proposed to account for the behavioral and neural dynamics observed in perceptual and motor decision-making tasks [Reddi and Carpenter, 2000, Smith and Ratcliff, 2004, Wong and Wang, 2006, Ditterich, 2006]. Although, slightly different in their mathematical descriptions most of them share in common the idea that decisions are the result of an internal process that accumulates noisy perceptual evidence in favor of each alternative [Gold and Shadlen, 2007]. Although these models have provided a great advance towards the understanding of the neural mechanisms involved in perceptual decision-making, they have mainly focused on highly constrained tasks that are performed in a controlled environment and it is not clear how well they would explain any other real-world decision making situation. One attempt to address this issue can be found in [Marcos et al., 2012], where a simulated robotic agent had to forage an open arena and perform action selection based on perceptual cues. Notably, the authors extended the Prefrontal competition to multiple choices and introduced a learning bias to the competitive decision-making dynamics. The present decision-making model shares with the one found in [Marcos et al., 2012] some basics dynamics, such as multiple choices competition, with the additional feature of being integrated with sub-cortical structures drawing the neural circuit partially responsible for goal-oriented behavior. For example, signals from the hypothalamic nuclei monitoring internal somatic states were further encoded by VTA as a control signal directed to decision-making modules. This interaction represents the motivation of the agent to act in order to obtain a reward and simulates the way dopaminergic signals might drive goal definition within the medial Prefrontal Cortex, defining the specific objective of foraging behavior [Lockie and Andrews, 2013, Luo et al., 2011]. This is consistent with the functions attributed to the Hypothalamus in determining the type of resources to be sought during foraging [Keen-Rhinehart et al., 2010].

One of the main aspects of operant conditioning can be found in the way consequences are linked to a given action. Reward mechanisms were taken into account by modeling some of the functions associated to the Basal Ganglia. In particular we focused on the object-value association found in the Striatum [Pennartz et al., 2011]. Here, reward associated with a given resource type was computed according to temporal discounting mechanisms such as those described in [Kobayashi and Schultz, 2008]. Reward signals mainly affected hypothalamic structures [Volkow et al., 2011] as the key sub-system for the definition of the needs of the agent. These signals were used to restore internal homeostatic levels responsible for the definition of low-level action pursuit. As discussed below, we did not take into account the effects of reward on motor control.

Another of the behavioral features observed in animal foraging and described by the operant conditioning paradigm is the optimization of behavior and maximization of reward [Abarca et al., 1985, Cabanac and Swiergiel, 1989, Lea and Tarpy, 1986]. Here, we have shown that the interaction of multiple subsystems organized according to the DAC principles can account for tendencies toward optimal foraging strategies found in the animal domain. Although a formal comparison with optimality theories in resource seeking [Charnov, 1976] is not present and outside the scope of this study (but see [Verschure et al., 2003b] for an analysis of optimality of DAC in Bayesian terms), we have measured a trial-by-trial increase of the performance of the agent.

We defined performance as the relationship between reward value and motor cost measured at every trial, which showed a highly significant inverse correlation. Such a measure however, was not conceived as an objective function to be maximized, rather it is seen as an emergent property of the behaving system. This apparent detail yields an important implication for the design of the system and for our understanding of the underlying neuro-computational principles, where distinctions between objective and emergent behavior plays a key role. Normative frameworks as, for instance, the free-energy principle [Friston, 2010, Friston et al., 2015] aim at explaining perception and behavior using a unified formulation that captures prediction and prediction error as both the objective and the consequences of animal behavior. In this sense goal-directed and uncertainty-reduction behaviors are merged in a unified probabilistic principle underlying the computational mechanisms of the animal brain. Therefore, the free energy principle seems to collapse the diversity of individual goals into the minimization of a single objective function referred as ”surprise”. In contrast, in DAC goals emerge from drives that emerge to satisfy different needs. In this sense, behavior is explained by a multitude of competing goals, where each goal can be decomposed in a set of sub-objectives distributed and optimized across the system (see [Herreros and Verschure, 2015] for a full commentary).

In support of this formulation, we suggest that the vertebrate brain requires diverse mechanisms of prediction and error reduction, which are not distinguished by the free energy principle. In particular, we have shown that a tendency towards optimization in our system can be identified at the subsystem level where each learning component of the system optimizes internal objective functions reducing a specific error. Within the cerebellar model, for example, optimization is defined as the minimization of the difference among the unconditioned and the conditioned response and it is captured by the decorrelation learning rule (described in Methods). Within the presented hippocampal model, optimality is defined as the trend to encode specific regions of an environment at higher resolution and stability through Hebbian learning. The interaction between grid and place cell populations led to an increase of place specific stabilization of place cells. Thus, as a learning and memory structure, the hippocampal region progressively optimizes the integration of spatial signals arriving from grid cell populations. In summary, the tendency towards optimization on multiple scales is not surprising and it is coherent with operant conditioning definitions where an animal is expected to achieve a diversity of sub-goals, from motor learning to self-localization and decision-making, that leads to the emergence of a behavioral pattern that optimizes the cost/reward relationship. The challenge, however, is to understand the sub-systems and the interactions that underly this macroscopic tendency.

Similarly, the behavioral optimization problem is the center of most of the reinforcement learning literature, which made an effort to formalize the outcome of a learning system according to a unified formulation based on the optimization of reward [Sutton and Barto, 1998] and going back to Thorondike's law of effect. In particular, several models tried to capture the behavior observed in classical (see [Balkenius and Morén, 1998] for a detailed comparison) and operant conditioning [Montague et al., 1995] with the aim to provide a descriptive explanation of conditioning phenomena. Whereas, multiple efforts were made to map such principles to the animal nervous system (see [Schultz, 2002] for a comprehensive review). These views greatly contributed to the understanding of reward-driven learning dynamics by elucidating the computational role of dopaminergic nuclei in the brain and their effect on behavior. Such a formulation however, describes the functions of specific neural circuits and must be necessarily integrated with the output of other subsystems that provide the building blocks for reinforcement learning computational mechanisms [Doya, 1999, Doya, 2000]. For instance, the acquisition of a state space, the definition of a value function, based on multiple goals defined by the intrinsic motivation, and its translation to a behavioral output can be interpreted as distributed elements described by the functions of complementary subsystems. Our results support this distributed view and suggest that a behavioral performance consistent with reinforcement learning formulations, namely the optimization of the reward-cost ratio, can emerge as a consequence of the interaction between localized modules even in absence of an explicit reinforcement learning rule. Therefore, the variability in the implementation of distinct systems raises the fundamental question of how this apparent conflict between the synthesis of a unified normative computational framework and the diversity of computational realizations in biology can be resolved. In addition, when normative frameworks describe macroscopic emergent behavioral properties, what is the leverage they give us in understanding the sub-systems that contribute to this performance? Further analysis is required to assess, from a theoretical perspective, whether the fitness of an agent, captured by optimization dynamics in the brain, is due to the role of specific localized structures serving as a 'monitoring system' (i.e. dopaminergic centers), or whether they are rather driven by more distributed objective functions.

Acquiring knowledge about reward and elaborate predictions for future action can be based on two learning methods. On one hand, in model-free learning the acquisition of value judgment of actions or events is the result of progressive experience. On the other hand, in model-based learning, representations of the environment and expectations are used to elaborate predictions on the future value of events. Despite this dichotomy, these two types of learning methods are not necessarily incompatible in the instrumental learning paradigm. Indeed, it has been argued that Pavlovian learning also involve a form of model-based evaluation [Dayan and Berridge, 2014]. In conjunctive model-based and model-free instrumental learning, both body and brain signals as well as the acquired model of the environment affect the computation of an event’s value and their motivational drive. Here we show that an agent acts based on both low-level motivational drives, such as hypothalamic signals, and learned spatial representation of the explored environment in order to optimize reward and minimize cost. Thus, our agent not only acquired a model of the environment based on its exploratory experience, but also used that model in order to evaluate future actions and value expectation. Further supporting the notion that the model-based and model-free dichotomy is not that useful in understanding adaptive behavior.

The use of a DAC-based system controlling artificial agents able to learn from the interaction with the environment is not new. The DAC architecture has been introduced 20 years ago. For instance, in [Verschure et al., 2003b] it was used to identify a novel non-neuronal feedback loop between perceptual and behavioral learning systems. Similarly, [Duff and Verschure, 2010] proposed a formal interpretation of such interaction as described in the correlative sub-space learning rule. Moreover, [Verschure and Voegtlin, 1998] proposed a study addressing the high-level computational functions underlying classical and operant conditioning in a mobile foraging robot. However, despite the novel insights these studies provided to the understanding of the organizational principles underpinning behavior, they did not account for a detailed mapping of the proposed computational strategies to the vertebrate nervous system. In contrast, multiple biologically constrained models of several individual brain areas, stemming from the theoretical account of DAC, have been recently proposed and validated against specific datasets. For example, [Herreros et al., 2013a] proposed a biologically plausible computational model of the cerebellar microcircuit used to investigate the role of the nucleo-olivary inhibitory pathways in adaptive motor learning. Renno-Costa [Rennó-Costa et al., 2010] successfully explained the mechanisms underlying rate-remapping dynamics in the Dentate Gyrus using a biologically valid computational model of the hippocampal formation. Moreover, [Marcos, Encarni Verschure, 2013] provided a model of the Pre-Frontal cortex that accounts for the modulation of neural dynamics in decision-making due to task memory. Finally, Guanella [Guanella et al., 2007] proposed a model of the Entorhinal cortex based on attractor dynamics that explains the self-organizing principles that govern space representation as encoded in the grid-like pattern of activation found in this area. Although, these models were up until now only validated in isolation, and did not pursue an integrated account of their functions within a single control architecture. Here, for the first time we proposed a cognitive control architecture for a real-world agent that combines the organizational principles proposed by DAC with the biologically valid computational principles found in several core areas of the mammalian brain, outlining their anatomical and functional interfaces and their role within a system-level neural and behavioral context. As such this could be considered an early example of an embodied real-world whole-brain model that we see as the key objective of theoretical neuroscience.

Outside of the DAC framework, multiple cognitive architectures have been proposed in order to solve a diversity of tasks. However, the coupling between a cognitive architecture and a theoretical framework is often silenced and many of the previously presented cognitive models reject specific organizational principles to be elaborated on. One example is the Soar architecture [Laird, 1987, Newell, 1994, Laird, 2012], which, as for DAC, aims at addressing a set of bio-inspired cognitive properties. However, Soar uses production rules to control its behavior, while the current version of DAC builds upon a biologically grounded framework where internal needs govern the embodied and situated agent's behavior. In that sense, the ACT-R architecture [Anderson, 1983, Anderson et al., 2004] refines the interpretation of the system in terms of cognitive neuroscience and extends the sub-symbolic processing in explicit perceptuo-motor modules. However, ACT-R takes a symbolic approach towards cognition, while DAC emphasizes embodiment and real sensory acquisition in order to act. SPAUN [Eliasmith et al., 2012] was recently presented as the largest model of the human brain that successfully managed to bridge the gap between complex behavior and complex neural activity. SPAUN was able to perform a number of tasks traditionally used in cognitive psychology, such as serial working memory or numerical sequence completion. Despite SPAUN's efficiency in solving rule-learning problems, it was tested through a disembodied agent performing within an abstract context, i.e. it did not have to solve the fundamental problems an embodied real-world agent has to solve. In contrast, we have implemented a cognitive architecture accounting for the functions of subcortical and cortical brain regions controlling an embodied agent acting within a real-world environment. In this sense, our approach is similar to the one proposed by the dynamic field theory [Sandamirskaya et al., 2013], where an embodied agent performs in a real-world setup solving tasks such as navigation and planning in dynamic environments. However, the dynamic field approach stands on a unitary computational principle, while we based DAC-X on a layered cognitive architecture that captures the diversity of the computational principles found in different structures of the mammalian brain. Our approach is also similar to [Prescott et al., 2006, Dasgupta et al., 2014] where a cognitive architecture including a model of the Basal Ganglia and the Cerebellum is used to solve a real-world foraging task in an open arena. Our strategy however differs in the number of computational elements used in the architecture and in the layered organizational principles used to build it. The concept of layered architectures is not new, and was already proposed in [Brooks, 1985] for controlling a variety of real-world agents. Although Brook's subsumption architecture wasn't originally conceived as a working hypothesis on the organizational principles of the brain and as such is neutral towards its interpretation, [Prescott et al., 1999] suggested an interesting parallelism among its layered layout and the structure of the vertebrate nervous system. In these regards the authors found in this work a plausible biological metaphor to understand the evolution of the neural substrate supporting animal behavior and cognition. Here, a key feature can be found in the way multiple controllers acquire a partial but complementary knowledge of the world that is integrated within a hierarchically organized architecture to originate behavior of a situated agent. The organizational principles of DAC however, show that the notion of well delineated modules does not hold, rather there are strong interdependencies between internal structures and layers supporting the bootstrapping of performance. Embodiment in particular represents a crucial aspect of DAC and it interprets the emergence of cognition as a bottom-up process originating from the physical interaction of an agent with its environment driven be dedicated constraints captures at the four layers of the DAC-X architecture.

Overall, despite the disparity of cognitive architectures that have been presented since [Newell et al., 1958], there has been great progress in the understanding of animal behavior and the human mind. For instance, ACT-R has made an effort to model human behavior, SOAR focused on the selection of suitable forms of knowledge, the Dynamic field theory architecture emphasized embodied cognition, and SPAUN has contributed to the understanding of high cognitive functions within fairly complex tasks.

Here we aimed to implement the neural circuitry responsible for operant and classical conditioning in the context of foraging. However, besides the achievements of DAC-X, we are still facing a myriad of experimental limitations. First, because of our focus on conditioning, many brain areas were not taken into account. Also, in our implementation, the functions of the addressed areas were modeled at different levels of description. Some areas were implemented in accord with anatomical and physiological constraints and learning principles described by neuroscientific literature, while functions of other areas were captured by more abstract yet biologically plausible implementations without considering their learning mechanisms or network dynamics. The Basal Ganglia, as an example of this last category, was modeled on its functions of reward delivery but the aspects of motor learning and action selection were not taken into account in detail.

Moreover, here as in previous DAC implementations, we have used the hoarding task benchmark so to match animal behavior. However, our static setup does not include the unpredictability and dynamics of real world situations. Furthermore, the exploratory arena used in our setup is very modest in terms of size if we compare it with trajectory lengths performed by rodents in real world tasks. In this sense, our task is a simplified version of the dynamics encountered in the real world. In addition, complex executive functions and reasoning were not addressed in our implementation, as they are in SOAR, ACT-R and SPAUN. However, none of these models have been instantiated in a real-world form yet. Finally, we have based our study on the dynamics presented by the H4W problem in order to solve a hoarding task. However, dealing with real world dynamics implies cooperating or competing with other coexisting agents. Moving into the H5W problem (adding a ”Who”) will imply computing additional processes such as social interaction and making predictions on the internal states of other agents.

In future work, we plan to elaborate the details of the neural models that were only captured in an abstract form, such as the Basal Ganglia, include a counter current attention model [Mathews et al., 2015] and expand the architecture to include a more complete connectomics based model of the Neo-cortex [Arsiwalla et al., 2015]. We will test this model towards more complex tasks including social interaction [Lallee et al., 2013], which we see as critically dependent on consciousness.

Part II

Chapter 2

ANTICIPATORY ACTIONS: ADAPTIVE MOTOR RESPONSES OR REACTIONS TO SENSORY PREDICTIONS?

In this chapter we focus on the feedback-feedforward cerebellar architecture presented in the previous chapter in the context of anticipatory control. We argue that, from a control perspective, the parallel organization implicit to standard cerebellar (inverse) models is not optimal when acquired feedforward predictions are violated. We propose an alternative control scheme that learns a feedforward prediction of a sensory event and test it in a simulated postural task where an artificial agent has to resist an external disturbance. We conclude that anticipating events in the perceptual domain, rather than advancing motor commands can achieve increased control robustness at lower effort. This chapter is based on: *Maffei, G., Sanchez-Fibla, M., Herreros, I., & Verschure, P. F. (2014). The role of a cerebellum-driven perceptual prediction within a robotic postural task. In International Conference on Simulation of Adaptive Behavior (pp. 76-87). Springer.*

2.1 Introduction

When learning to perform a skillful motor task such as skiing, one has to be able to resist perturbations issuing appropriate motor actions in order to maintain constant balance and equilibrium and avoid falls. The novice skier would tend to correct a disturbance, due to wind or irregularity of the slope, at the time it is experienced. However with practice and exercise, he would be able to recognize an incoming disturbance and to trigger a preparatory motor action in order to minimize the effect of the perturbation on his body configuration, his balance and ultimately on his performance. Postural adjustments are described within the realm of motor control as small muscular responses which constantly adjust the body configuration to maintain barycenter position and equilibrium while walking, lifting objects or during collisions with obstacles [Massion, 1994, Peterka, 2002].

Experimental studies found that healthy subjects involved in motor tasks, such as catching a ball [Shiratori and Latash, 2001], or lifting objects of different weights [Flanagan et al., 2001], rely on progressively acquired motor patterns that enhance performance. Electromyographic recordings show that compensatory muscular activity posterior to the experienced perturbation is increasingly coupled with preparatory responses, possibly driven by the adaptation to the disturbance. Such results suggest that a preparatory action can be learned and initiated in advance, and that postural adjustments can be decomposed in two elements: compensation and anticipation. Similar studies conducted on cerebellar patients show that they lack predictive anticipatory actions and correct response magnitude scaling when tested in postural tasks, such as standing still on a sliding platform [Timmann and Horak, 2001b, Horak et al., 1994] or minimizing the arm vertical shift while catching a falling object [Lang and Bastian, 1999, Serrien and Wiesendanger, 1999a]. These findings would make the cerebellum an ideal candidate as neural substrate involved in the acquisition of adaptive postural motor responses.

Despite several mechanisms underlying postural control have been widely studied, it remains unclear what strategy is adopted by the brain in order to issue predictive motor responses. Moreover, it is poorly understood how anticipatory and compensatory components are integrated into a single response. A possible explanation is that these responses are both the result of an association between a sensory signal and a motor response. The former would be triggered by a sensory signal co-occurring with the perturbation, while the latter would be initiated by a sensory signal preceding the perturbation. The total motor response would therefore be the sum of the two components separately acquired and combined in more peripheral areas such as the spinal cord [Bizzi et al., 2008]. However results on the topographic organization of sensory and motor representations of the hand in the human cerebellum show that, unlike the neocortex, sensory and motor patches for the same finger do not overlap systematically, but are closely interdigitated in a nearly unrelated fashion [Wiestler et al., 2011]. The suggested close interaction between sensory and motor cerebellar circuits leads to an alternative explanation. It is indeed possible that compensatory and anticipatory responses are the result of the interaction of two predictions of different nature. Compensation could be achieved by mapping a sensory input into an adaptive motor response. Differently, anticipation could be achieved by associating a sensory signal anticipating the perturbation with the sensory signal perceived at the moment of the perturbation. This sensory prediction would then trigger the compensatory action in an anticipatory way.

With these hypotheses in mind we propose an adaptive control architecture formed by a compensatory and an anticipatory layer. The former acts as a fast feed-forward controller that corrects the effect of a perturbation after it has been experienced. The latter is responsible for anticipating the incoming perturbation and initiating an action in advance. We compare two alternative hypotheses to test the nature of anticipation: 1) a sensory to motor prediction, which associates a sensory event preceding the perturbation with an adaptive motor response, and 2) a sensory to sensory prediction, which associates a sensory signal with the expected sensory outcome of the perturbing event and, in turn, triggers a motor response. In addition, we explore the role of cerebellar adaptive properties, proposing a learning strategy based on a model of the cerebellum where both compensatory and anticipatory components can be acquired by two independent instances of the same cerebellar controller [Herreros et al., 2013b, Herreros and Verschure, 2013b, Maffei et al., 2013] .

Coherently with human equilibrium tasks [Horak et al., 1994], we devise a simulated cart-pole setup in which a robot has to minimize the error provoked by a perturbation directed to the pole. Similarly to [Maffei et al., 2013], where a real robot had to maintain constant speed anticipating the effect of the collision with an obstacle, the agent is equipped with sensors that allow to measure impact force and proximity. The former provides a sensory signal at the moment of the impact with an object allowing fast compensation. The latter perceives the distance from the object allowing anticipation. We propose an experimental procedure to compare the two architectures in terms of learning curve, error minimization and motor cost. Results suggest that an agent relying on an anticipatory sensory prediction can remarkably reduce the error with less effort. Moreover, this architecture appears to be more robust in case of ambiguity of the stimulus, as shown in trials where a sensory signal previously associated with a perturbation is no more reliable. Finally we discuss the obtained results in the light of recent physiological and behavioral evidences supporting the versatility of the cerebellum in learning associations outside the scope of motor control. Implications for bio-mimetic robot control are also taken into account.

2.2 Methods

2.2.1 Setup

In order to study the possible role of cerebellum in anticipatory responses to postural perturbations we devise a simulated physics based setup implementing the cart-pole dynamics (fig. 2.1, $Left$). A simulated agent has to resist a postural perturbation through anticipation in order to minimize error and energy cost. The agent is able to slide on a horizontal surface controlling one degree of freedom with the goal of maintaining the pendulum in a constant vertical equilibrium performing control against force of gravity and external perturbing forces. The agent is equipped with a proximity sensor and a pressure sensor detecting the distance to external bodies and eventually the magnitude of the force produced by the collision with them. A colliding object is directed to the extremity of the pendulum with a given constant velocity and force, therefore provoking a perturbation that affects the pendulum position and the state of equilibrium of the agent. The goal of the agent is to learn to associate sensory inputs to finely tuned motor responses in order to firstly compensate and secondly anticipate the perturbing event both in terms of magnitude and timing, therefore minimizing the pendulum deviation from the state of equilibrium with the minimum effort.

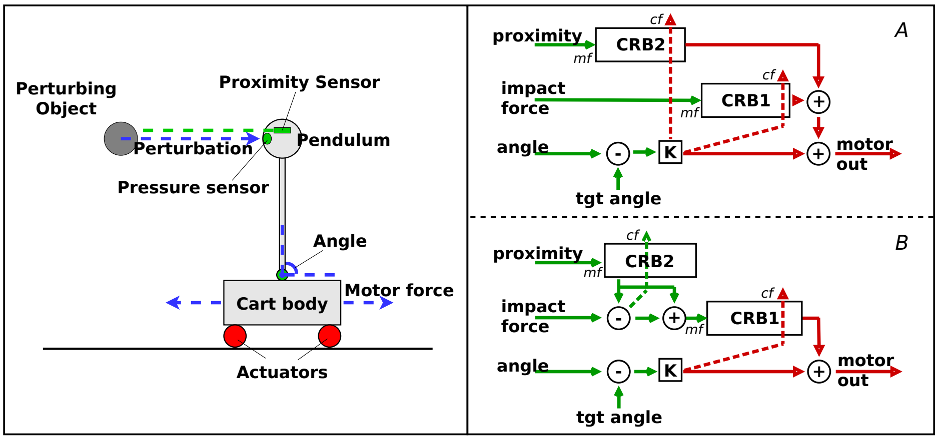


Figure 2.1: $Left$. Cart-pole setup. Cart represented in equilibrium position (pendulum at 90 degrees) and incoming perturbation. Sensors are represented in green, actuators in red. $Right$. Computational Architectures. A. Sensory-to-motor and B. Sensory-to-sensory hypotheses. Sensory signals are represented in green. Motor signals are represented in red. Adaptive components are represented by two cerebellar modules ($CRB1$ and $CRB2$ respectively). Note the different input-output configuration for $CRB2$ in A, where the proximity signal is associated to a motor response and B, where the proximity signal is associated to the impact force. Mossy fibers (mf) represent the microcircuit input stage, while Climbing fibers (cf) provide the teaching signal (dashed line) via Inferior Olive (not displayed).

K represents the gain of the reactive proportional controller that converts the sensory input into motor output.

2.2.2 Learning algorithm.

The bio-mimetic learning algorithm at the core of the behavior of the agent is based on an analysis-synthesis adaptive filter implementation mimicking the learning strategy of the cerebellar microcircuit [Dean et al., 2010b, Herreros and Verschure, 2013b]. The cue signal, representing the input conveyed by mossy fibers, is decomposed into several signals mimicking the expansion of information into cortical basis occurring within the cerebellar granular layer. The signal of the cortical basis is generated producing a fast excitatory component and a slow inhibitory one. Each component consists of a double exponential convolution with time constants randomly drawn from two flat probability distributions (a fast time constant, ranging from 5 to 50 ms and a slow one ranging from 50 ms to 2.5 s controlling the raise and the decay of the basis respectively, coherently with the physiological range of the time constants of the slow currents in the granular layer [van Dorp and De Zeeuw, 2014]). The value obtained after the two convolutions is then thresholded and scaled for each basis.

The output of the cerebellar controller is given by: $CR(t) = [\mathbf{p}(t)^T \mathbf{w}(t)]$ where $\mathbf{w}(t)$ is the vector of weights and $\mathbf{p}(t)$, the vector of basis, both in column form.

The weights are updated using the de-correlation learning rule: $\Delta w_j(t) = \beta\ E(t)\ p_j(t - \delta)$ where β is the learning rate and $E(t)$ is the error signal, computed by the inferior olive output. δ provides the latency of the nucleo-olivary inhibition. The value of δ determines how much the adaptive action anticipates the reactive one, and how much it has to exceed the feedback delay [Miall et al., 1993].

Finally, the error signal for the cerebellar system is computed as the difference between the scaled cerebellar output and the unconditioned stimulus (US) signal as follows: $E(t) = US(t) - k_{noi}CR(t - \delta)$

Computational Architecture. The agent implements a control architecture composed by three modular layers (fig. 2.1, $Right$, A, B).

The first layer implements a feedback reactive controller which computes the difference between a given target angle (equilibrium point at 90 degrees with respect to the horizontal axis) and the actual angular position of the pole. The error, multiplied by a gain, is mapped into a reactive motor response which moves the cart accordingly, therefore readjusting the position of the pendulum with respect to the target.

The second layer implements an instance of the cerebellar microcircuit, and it is responsible for acquiring compensatory responses. The cue signal is given by the force input resulting from the collision with the perturbing object. The signal to be learned ($teaching\ signal$) is given by the output of the reactive controller, encoding the action necessary to compensate for the pendulum error. The output of the controller is an acquired compensatory motor response acting in a feed-forward manner and summing to the output of the reactive controller.

The third layer of control, implementing a second instance of the same cerebellar microcircuit, is responsible for anticipatory responses.

In order to study how anticipation is performed within the context of anticipatory postural responses we propose two possible configurations which reflect, at the implementation level, the alternative hypotheses on the nature of such responses.

On one hand we propose a sensory-to-motor ($S2M$) configuration (fig. 2.1, $Right$,A) which takes the input from the proximity sensors as cue signal and the output of the reactive controller as error signal. The output is represented by a feed-forward anticipatory motor response which is summed to the motor response of the compensatory and reactive layer.

Alternatively, we test a sensory-to-sensory ($S2S$) configuration (fig. 2.1, $Right$,B). In this case the cue signal is given by the proximity sensor value while the teaching signal is given by the force input resulting from the collision with the perturbing object. This controller outputs a predicted sensory signal anticipating the sensory consequences of the collision. The prediction is subtracted with a small delay from the real incoming sensory signal coherently with neurophysiological data on sensory integration between cerebellar driven prediction and actual somatosensory feedback [Blakemore et al., 1998]. The net force signal finally inputs the compensatory controller, which in turn triggers an action preceding the perturbation, therefore producing both an anticipatory and a compensatory motor output.

2.2.3 Experimental Design.

The experimental session proceeds on a trial by trial base, having the agent set at a given position in a state of equilibrium at the beginning of every trial.

During each trial (5 seconds duration), a colliding object (25 kg) is directed to the extremity of the pendulum with a velocity equals to 1 m/s, therefore provoking a perturbation of 25 N affecting the pendulum position and the state of equilibrium of the agent.

We run a set of experiments to primarily test the effectiveness of such layered architecture in associating sensory inputs to finely tuned motor responses. The goal is learning to predict the perturbing event both in terms of magnitude and timing, issuing a motor action that minimizes the pendulum deviation from the state of equilibrium with the minimum effort.

We therefore compare the learning performances of the two alternative anticipatory configurations both in terms of learning capabilities and robustness to events that violate the acquired associations. In the first experiment we run a session of 50 trial for each of the two proposed architectures with the goal of comparing the dynamics in the acquisition of the compensatory and anticipatory responses. We are particularly interested in error minimization and efficiency in motor action cost. In the second experiment we test the acquired responses under a condition in which the perturbing object is still activating the proximity signal but is not provoking a perturbation anymore. We test both architectures under this condition for 10 trials looking at the robustness and flexibility of the architecture when events violate the acquired prediction.

2.3 Results

2.3.1 Experiment 1.

The goal of this experiment is to test the learning capabilities of the proposed computational architecture under $S2M$ and $S2S$ conditions.

We train the agent to co-acquire both a compensatory and an anticipatory response with the goal of minimizing the deviation of the pendulum from a given target angle. The training session lasts 50 trials for both S2M and S2S architectures, during which a perturbation of constant magnitude of 5 N is sent to the pendulum with a delay of 1 sec from the beginning of the trial.

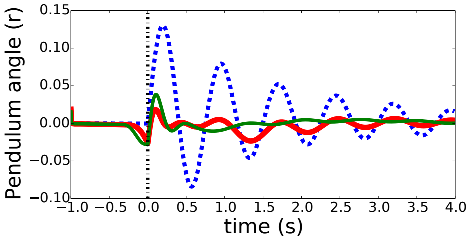


Figure 2.2: $Left$. Pendulum Angle. $Right$. Normalized mean cumulative error. $Colorcode$: Blue dashed: reactive controller only (mean 10 trials). Red thick: S2M trained controller (mean 10 trials). Green thin: S2S trained controller (mean 10 trials). Black: perturbation onset.

This perturbation magnitude provokes the sole feedback controller to reactively adjust the position of the pendulum (fig. 2.2, $Left$) showing a prominent oscillatory pattern which reduces the error over time but never stabilizes the pendulum to the desired position during the whole duration of the trial.

Under both $S2M$ and $S2S$ conditions, at the end of the training the robot is able to issue both an anticipatory and a compensatory response that minimize the effect of the incoming perturbation on the position of the pendulum (fig. 2.2, $Left$). We notice that a small deviation from the target is introduced by the agent itself as a consequence of the anticipatory response approximately 0.3 seconds before the perturbation. This well timed response allows the minimization of the perturbing force provoked by the collision, and therefore stabilizes the normalized cumulative error around a value of 0.05 (fig. 2.2, $Right$). If compared with the early trials, the intervention of the reactive controller at the end of the session is greatly reduced (fig. 2.4, $Right$) minimizing therefore the amount of energy required to stabilize the pendulum in a correct position.

The adaptive motor response at the end of the training experiment can be decomposed into two different elements.

Under the $S2M$ (fig. 2.3, $Left$) condition we observe that a motor response triggered by the sensed impact with the colliding object is issued at the moment of the collision.

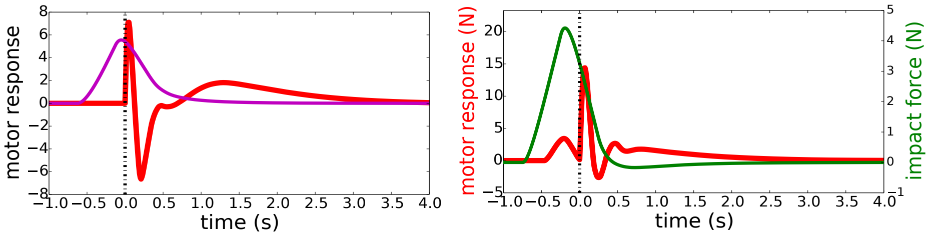


Figure 2.3: $Left.$ S2M adaptive responses. Magenta thin: anticipatory motor response. Red thick: compensatory motor response. Black dashed: perturbation onset. $Right.$ S2S adaptive responses. Green thin: anticipatory sensory prediction. Red thick: anticipatory and compensatory motor response. Black dashed: perturbation onset.

It peaks before the response of the sole reactive controller, and allows a faster compensation. An additional motor response triggered by the proximity signal is issued before the perturbation onset, and it is added to the total motor response allowing anticipation.

Under the $S2S$ condition we notice that a single adaptive motor response is issued accounting for both compensation and anticipation (fig. 2.3, $Right$). One single motor response is acquired at the level of the compensatory controller where, similarly to the $S2M$ architecture, a force sensory input signaling the impact with the perturbing object triggers a fast motor response. However the anticipatory response is achieved by predicting the force sensory input from the proximity signal in a sensory to sensory fashion. The predicted signal then inputs the compensatory module which triggers a motor response in an anticipatory way.

Both architectures perform almost equally, with similar performance in terms of error minimization and learning curve slope (fig. 2.4, $Left$). However the total adaptive energy produced by the $S2M$ architecture is significantly higher than the one produced by the $S2S$ architecture (fig. 2.4, $Right$).

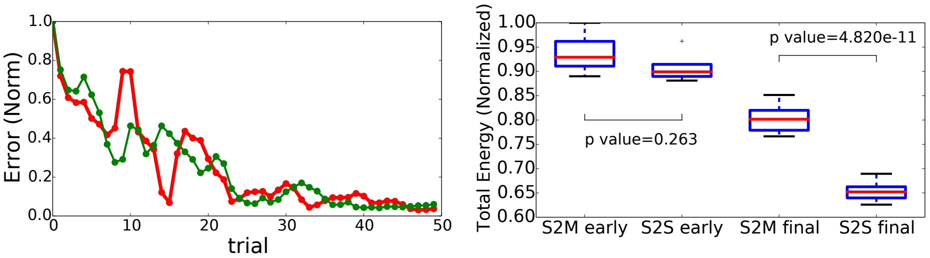


Figure 2.4: $Left$. Learning curve during the training session (normalized cumulative error). Red thick: sensory-to-motor architecture. Green thin: sensory-to-sensory architecture $Right$. Total motor energy (sample of 10 trials per condition). We compare significance between S2M and S2S architectures at early learning stage and advanced learning stage.

2.3.2 Experiment 2.

The goal of the second experiment is to test the reliability and flexibility of the architecture in case of unexpected conditions, namely the amount of self induced error in the case that a cue signal previously anticipating the collision is providing no perturbation.

Once both anticipatory and compensatory responses are acquired we run a short session of 10 trials for each architecture setting the mass of the colliding object to 0g, and therefore producing a null perturbation.

Both controllers introduce an error triggered by an erroneous anticipatory response. However we observe that the self-induced error in the case of the S2M architecture is significantly higher and varying than the one introduced by the S2S architecture, which in turn appears to be more stable and more resistant to unexpected conditions (fig. 2.5, $Right$).

This difference could be due to the comparison between predicted sensory signal and real sensory signal found in the sensory to sensory architecture (fig. 2.1, $Right$).

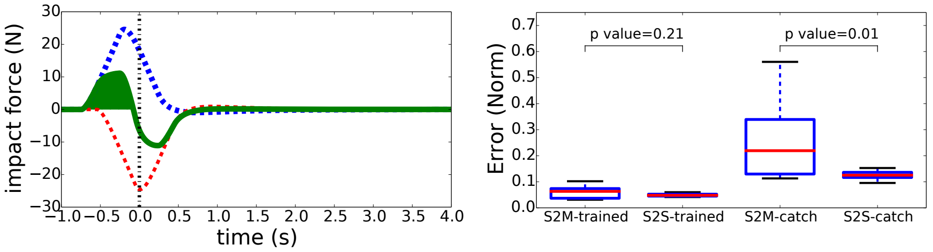


Figure 2.5: $Left$. Sensory prediction - sensory input mismatch in catch trial. Blue dashed thick: sensory prediction output of $CRB2$. Red dashed thin: sensory prediction error. Green solid: net sensory input to $CRB1$. Filled green area: positive part of net sensory input triggering a partial motor response. $Right$. Normalized cumulative error (sample of 10 trials per condition). We compare significance between S2M and S2S architectures at advanced learning stage and during catch trials (experiment 2).

Here the predicted sensory input is subtracted from the real incoming sensory signal with a delay of 200 ms. The resulting net signal inputting the first cerebellar module rises according to the erroneous prediction and subsequently decays to negative values as a result of the mismatch between expected and real perturbation (fig. 2.5, $Left$). Such signal triggers just a partial adaptive motor response, avoiding the full execution of the action and therefore introducing less error and less variability to the motor response.

2.4 Discussion

Within this study we are interested in the computational mechanisms underlying the acquisition of anticipatory responses in a postural task. We compare two alternative architectures representing two different hypotheses: anticipation either as an acquired sensory-to-motor association or as an acquired sensory-to-sensory association. We also propose that cerebellum could be the neural substrate responsible for the acquisition of both anticipatory and compensatory responses.

Both hypotheses would be consistent with cerebellar learning mechanisms. Avoidance learning studies [Jirenhed et al., 2007, Hesslow, 1994] indeed show how animals learn to produce a predictive motor action triggered by a sensory cue in order to avoid a noxious stimulus. This view suggests that the output of purkinje cells, the sole output of the cerebellum, would be directly contributing to motor responses. This would therefore be coherent with the sensory-motor association hypothesis which has been shown capable of efficiently perform in terms of error minimization within a postural task. According to this view the integration of two independent motor responses might take place at a peripheral stage, such as the spinal cord [Bizzi et al., 2008]. However the overall performance of the sensory-motor controller is not completely satisfying in terms of energy costs and robustness to unexpected conditions.

Postural control can be considered a complex task involving several sensory-motor interactions [Massion, 1994]. We suggest that with increasing complexity of the task, more sophisticated learning strategies might need to be applied by the brain. Interestingly, recent studies on the non-motor functions of the cerebellum [Ramnani, 2006c] suggest that it would be capable of predictions outside the scope of adaptive motor control. Neuroimaging studies have shown that the human cerebellum is active during somatosensory processing [Gao et al., 1996] as well as visual and auditory perceptual tasks [Baumann and Mattingley, 2010]. It has been also shown that cerebellar patients perform poorly in pure perceptual associative tasks where the prediction of a sensory signal is required to be learned from a second sensory signal [Roth et al., 2013b, O'Reilly et al., 2008]. Finally, anatomical studies show that the cerebellum has distinct projections to brain areas important for perception [Dum and Strick, 2003].

These findings could support the hypothesis which considers anticipatory responses in the brain as the result of the interaction among a pure sensory-to-sensory prediction and a sensory-to-motor prediction. This view would not contradict the well established findings on the cerebellar direct contribution to motor control, but would extend its adaptive properties to a more perceptual domain. We hypothesize that a possible advantage of relying on sensory predictions is the tendency to minimize the amount of action performed to achieve some motor goal. As we show in our results the sensory-to-sensory architecture achieves error minimization in a more efficient way compared with the sensory-to-motor one. Our assumption is that learning to produce one adaptive response driven by an expected sensory signal requires less energy than producing two, eventually antagonist, motor responses with different temporal profiles.

This may not be the only advantage. We show indeed that the sensory-to-sensory architecture is more stable and robust within conditions in which an ambiguous stimulus is suddenly provided. This result could be incongruous with findings from conditioning studies and avoidance learning. According to this paradigm a subject tested on catch trials would trigger a full motor response as result of a sensory-motor association, where diminished response would be gradually expected during the extinction phase. This view would therefore be more congruent with the sensory-motor hypothesis.

However behavioral results have suggested that healthy subjects trained to resist an expected perturbation tend to minimize self induced error in case of unexpected conditions, reducing the acquired motor response since the first trial [Crevecoeur and Scott, 2013]. This view is incongruous with the previous one, possibly requiring a more complex explanation. We propose that relying on a sensory prediction for anticipation could have an advantage in terms of performance, and eventually safety for an individual. As also proposed in [Crevecoeur and Scott, 2013], this could indeed be a mechanism that weights a sensory expectation with an actual sensory signal, partially preventing ambiguous stimuli to trigger inappropriate motor responses. The direct interaction between sensory and motor predictions in the cerebellum could be supported by physiological data showing a scattered an interdigitated topographic organization of sensory and motor areas in the cerebellar cortex [Wiestler et al., 2011]. As suggested by the authors such an arrangement may enable the cerebellum to quickly form new, and often context dependent, sensory-motor associations . This would ultimately be an important computational feature for learning new motor tasks, in which sensation and action might take on novel relationships.

The advantages above described could finally benefit robotic architectures. The proposed bio-mimetic approach would allow a more efficient adaptive control of posture in humanoid robots and, in general, a minimization of errors during navigation and manipulation tasks. The importance of learning to anticipate, as found in humans and animals, can therefore be directly applied to agents able to learn useful sensory-motor contingencies from the interaction with the environment. The proposed learning strategy represents indeed a model-free approach where physical properties of the environment are not assumed but progressively acquired. This could imply more flexibility and ability for an agent in adapting to its surroundings and learning appropriate motor responses from experience.

Chapter 3

THE PERCEPTUAL SHAPING OF ANTICIPATORY ACTIONS

This chapter builds on the control insights from the previous chapter and proposes a new theoretical framework where the standard parallel cerebellar architecture for anticipatory motor control (FEL) can be recast into a hierarchical scheme that reflects the causal structure of perceptual events in the environment. Contrary to standard cerebellar schemes, this novel approach can achieve anticipation while solving the control problems of violation and generalization in a simulated biologically plausible postural task, and it exhibits a performance consistent with behavioral literature. This theoretical framework has implications for cerebellar theory as it proposes that hierarchical descending predictions and ascending prediction errors can be implemented in cortico-cerebellar loops that only modulate behavior at the last stage of the control chain, as also proposed by normative accounts such as active inference.

This chapter is based on: *Maffei, G., Herreros, I., Sanchez-Fibla, M., Friston, K. J., & Verschure, P. F. (2017). The perceptual shaping of anticipatory actions. In Proc. R. Soc. B (Vol. 284, No. 1869, p. 20171780). The Royal Society.*

3.1 Introduction

Anticipatory motor actions, thought to depend on the cerebellum [Massion, 1992, Kolb et al., 2004, Timmann and Horak, 2001a], are part of our everyday behavior: from walking [MacKinnon et al., 2007, Xu et al., 2004], to grasping [Lacquaniti and Maioli, 1989, Lang and Bastian, 1999, Cordo and Nashner, 1982] and to riding a bicycle [Vansteenkiste et al., 2014]. The question then arises as to how these actions are controlled? Decades of research in motor control support the notion that internal models are key to skillful performance [Kawato, 1999, Shadmehr et al., 2010b, Wolpert et al., 2011]. Specifically, this research has highlighted two kinds of internal models: forward models, which map the efference copies of motor commands into their expected sensory consequences [Miall et al., 1993, Wolpert et al., 1998b]; and inverse models, which map desired sensory outcomes into their required motor commands [Kawato et al., 1987, Kawato and Gomi, 1992].

However, here we argue that offering an alternative to these interpretations is a pressing issue for the field of motor control as neither forward nor inverse models (in their standard formulation) can explain the versatile anticipatory control observed in animals. In particular, standard forward models allow for rapid feedback control in the presence of the long transport latencies of the nervous system [Miall et al., 1993] or action planning [Shadmehr et al., 2010b] but, as they exclusively predict the consequences of motor commands, they cannot anticipate disturbances that are not contingent upon those motor commands [Crevecoeur and Scott, 2013]. That is, one cannot call upon efference-driven forward models to support behaviors that precede external events. This obvious limitation has led researchers to conclude that preparatory actions should result from inverse models that output anticipatory motor signals [Albert and Shadmehr, 2016, Ruan et al., 2007, Sainburg et al., 1999, Shibata and Schaal, 2001, Thoroughman and Shadmehr, 2000]. The benchmark computational model for that theory is feedback error learning (FEL), which offers both an adaptive motor control architecture [Gomi and Kawato, 1993] and a theory of cerebellar function [Kawato, 1999, Kawato and Gomi, 1992]. In

FEL, predictive actions are the result of anticipatory motor signals, learned by shifting forward in time the output of the feedback controller [Albert and Shadmehr, 2016, Thoroughman and Shadmehr, 2000]. However, we will show that inverse model schemes present some important limitations in the context of anticipatory control. For instance, while rapid corrections of erroneous anticipatory actions are commonly reported in biological systems, most notably in experiments that include catch trials (i.e. trials where a predictable disturbance is signaled but not delivered) [Barnes et al., 2000, Witney et al., 1999], FEL has no mechanism to correct feed-forward motor responses once the course of events violates a prediction. In addition, FEL acquires motor commands that are tied to the dynamics of the plant that it controls and cannot easily be generalized to new configurations. However, experimental evidence suggests that in humans, anticipatory responses are still effective even if one changes the posture and/or the effector after learning [Criscimagna-Hemminger et al., 2003, Ahmed and Wolpert, 2009]. Hence, given that standard forward and inverse models cannot fully account for anticipatory control, alternatives should be considered that both overcome the theoretical and practical limitations of these motor-centric accounts and resolve the forward-inverse model dichotomy.

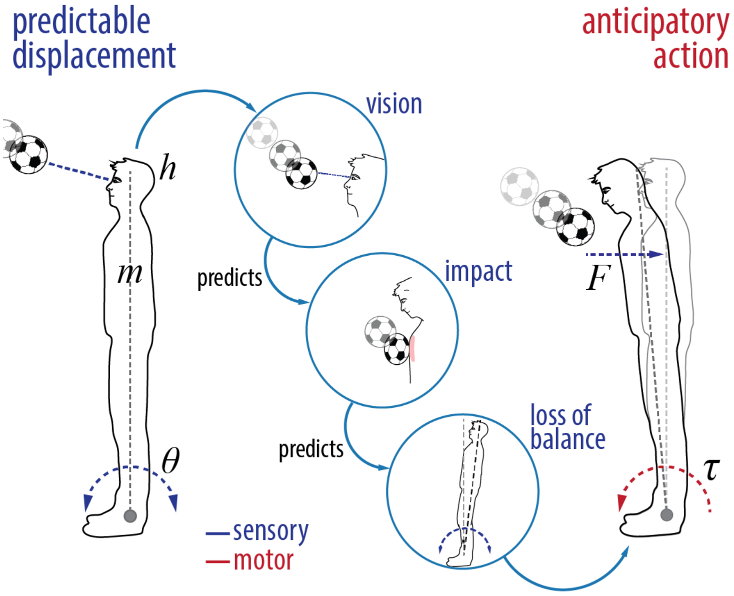


Figure 3.1: **Conceptualization of the Hierarchical Sensory Predictive Control (HSPC) hypothesis.** A predictable displacement caused by a soccer ball directed to the chest elicits an anticipatory response that minimizes the loss of balance before it is perceived. In HSPC, the anticipatory response is the result of a hierarchy of descending sensory predictions from distal (visual detection) to proprioceptive (impact) to vestibular (loss of balance) modalities, where each modality advances in time the expected consequences on the next modality until the predicted error in balance triggers a reflexive action in a feed-forward manner. The minimal model for this behavior is an inverted pendulum of mass (m) and height (h), whose error in angle (θ) is minimized by generating a torque (τ) at the ankles that counteracts the disturbance (F).

Here, we advance the hypothesis that biological anticipatory control can be explained by the ability of the brain to advance predictions of future perceptual events [Roth et al., 2013a] and use those predictions to drive the motor system in an anticipatory way [Herreros et al., 2016]. We formulate this hypothesis in computational terms by proposing the cerebellar-based

Hierarchical Sensory Predictive Control (HSPC) architecture, in which internal models issue sensory predictions that facilitate anticipatory control, with motor signals (i.e. efference copies of motor commands) playing no role in adaptation itself. With that, HSPC challenges the inverse model interpretation of anticipatory control - and, indirectly, the ’motor-centric’ forward-inverse model dichotomy. More precisely, we suggest that, in contrast to the FEL hypothesis, where predictive actions are the result of anticipatory motor signals, anticipatory actions can be controlled by predictive sensory signals, becoming reactions to events that are brought forward in time [McIntyre et al., 1995, Modayil and Sutton, 2014, Serrien and Wiesendanger, 1999b]. Moreover, in HSPC the internal generation of sensory predictions can mirror the (hierarchical) causal structure of the sequence of perceptual events (fig. 3.1). HSPC builds on the hypothesis that motor control can be understood as a process of sensory-sensory learning where sensory predictions are only mapped onto motor commands at the late stage before motor execution, for example through reflexes, as proposed in the Distributed Adaptive Control (DAC) theory and formalized in the theory of Active Inference [Adams et al., 2013, Verschure et al., 2003a, Friston, 2011a]. At the theoretical level, this hypothesis has been studied mostly within the active inference framework, using generative hierarchical models and focusing on the aspect of reformulating control as Bayesian inference [Friston, 2011a, Friston et al., 2010], whereas DAC generalized it to robot-based foraging tasks showing Bayesian equivalence [Verschure et al., 2003a]. Hence, here we propose for the first time a detailed computational and practical treatment of the sensory-sensory learning hypothesis in the context of anticipatory actions. To this end, we provide a systematic comparison between HSPC and FEL by synthesizing each hypothesis into an architecture applied to a postural control task, minimally modeled as the stabilization of an inverted pendulum through a torque at its base (i.e. ankles; fig. 3.1), demonstrating how learning in the sensory rather than in the motor domain can account for the robustness and generalization capabilities of biological control systems with emphasis on the relation between the cerebellum and the neo-cortex. In summary, this study presents an approach to motor control that could provide an alternative interpretation of the physiology of anticipatory control and contribute to the theory of cerebellar learning.

3.2 Methods

In order to compare the behavior of a control strategy based on motor anticipation (FEL) with one based on sensory prediction (HSPC), we synthesize these hypotheses into two architectures that control an inverted pendulum (a common model for bipedal postural control - see [Gage et al., 2004] for review; fig. 3.2-B,C; electronic supplementary material, fig. S1) engaged in an Anticipatory Postural Adjustment (APA) task. This task, in line with experimental psychology paradigms [Kolb et al., 2004, Santos et al., 2010b] (fig. 3.2-A), requires the agent to learn an appropriate combination of anticipatory and compensatory responses to minimize the effect of a disturbance (i.e. loss of balance) signaled by a cue.

3.2.1 Model of the Agent

The inverted pendulum actuated by a torque (τ) at its base is modeled as follows:

$$mh^2\ddot{\theta} = mgh\sin(\theta) + \tau + Fh\cos(\theta)$$

(1)

The pendulum has a mass (m) of 67 Kg and a height of its center of mass (h) equal to 0.85 m. θ measures the angular deviation from the vertical position. The disturbance is introduced as a force (F) parallel to the ground applied to the center of mass.

3.2.2 Control Architectures

The APA task involves three different sensory modalities: distal (perceiving a cue that precedes the collision), proximal or proprioceptive (sensing the

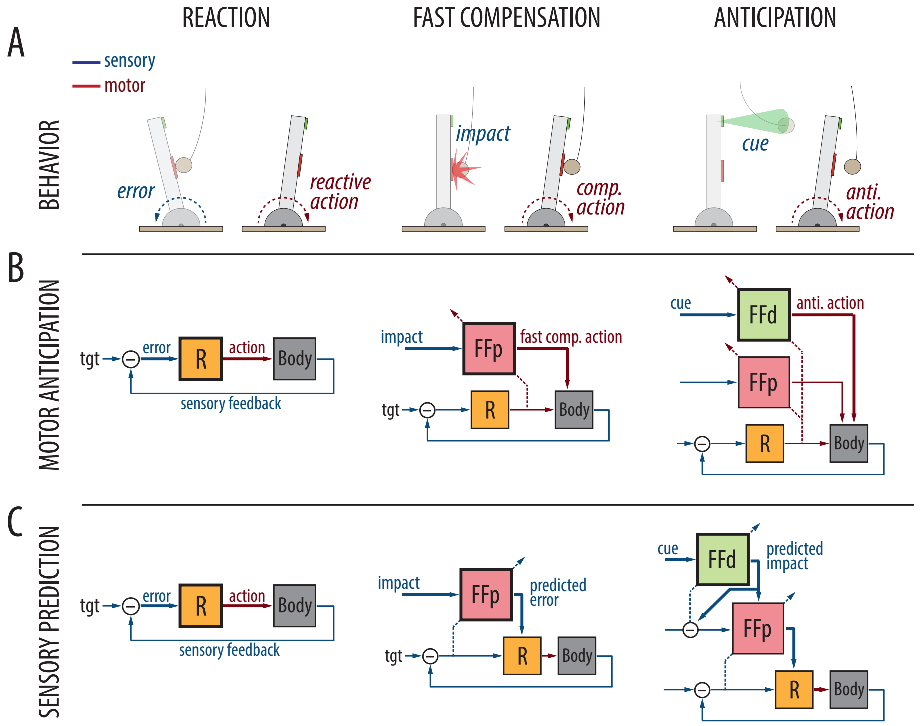


Figure 3.2: **Motor anticipation (FEL) versus sensory prediction (HSPC) strategies.**(a) Different responses are elicited by different sensory modalities. (i) A corrective reaction is triggered by the perceived postural error. (ii) A fast compensatory corrective action is triggered by the perceived impact (proximal stimulus). (iii) An anticipatory action is triggered by the distance to the obstacle (distal stimulus). (b) Motor anticipation strategy (FEL). (i) A postural error is converted into a reflexive action by a feedback controller (R). (ii) A feed-forward compensatory action associated with the impact signal is acquired by the proximal adaptive module (FFp) on the basis of the feedback response to the error. (iii) A feed-forward anticipatory action associated with the distal cue is acquired by the cerebellar distal module (FFd) on the basis of the same feedback response. (c) Sensory prediction strategy (HSPC). (i) Reflexive action elicited as in FEL. (ii) Feed-forward compensatory action: triggered by

the proximal cue and learned from the closed-loop error, a counterfactual error is issued by the proximal module (FFp) in response to the proximal cue driving the feedback controller. (iii) Anticipatory action: evoked by the cue, a prediction of the expected impact issued by the distal module (FFd) triggers the compensatory response in an anticipatory manner.

magnitude of the impact on the body) and vestibular (sensing the postural effects of the impact, i.e. the inclination). Each modality enables a different type of response: distal sensing allows for preparatory responses, proximal sensing for fast compensation and vestibular for compensation through feedback control [Lacquaniti and Maioli, 1989, McIntyre et al., 1995, Mohapatra et al., 2012, Peterka, 2002, Savelsbergh et al., 1992] (fig. 3.2-A). Note that a similar distinction between distal and proximal sensory modalities can be found in [Verschure et al., 2003a, Friston, 2011b] to account for sensory predictions within extrinsic and intrinsic frames of reference, respectively.

Feedback controller

The agent is stabilized by a torque generated through a proportional-derivative feedback controller as follows:

$$\tau_{fb}(t) = -k_p \theta(t - \delta_s) - k_d \dot{\theta}(t - \delta_s)$$

(2)

Note that in the error term we use the angle and angular velocity values delayed by δs (=100 ms) to account for the latency of the error feedback.

Adaptive feed-forward modules

In addition to a feedback controller, both architectures include the same adaptive feed-forward modules to process the proximal and distal cues. That adaptive feed-forward module (i.e. inversion of a forward or generative model) is implemented as an adaptive filter extended with an eligibility trace mechanism [Herreros and Verschure, 2013a, Herreros et al., 2013a, Maffei et al., 2014]. Each feed-forward module receives a single sensory input signal that is expanded into N (=20) different signals or bases. Each basis corresponds to the convolution of the (sensory) input with an α signal that can be formulated as two serially linked leaky integrators with identical time constants. For a particular basis, its output value is generated as follows:

$$z_j(t+\Delta t) = \gamma_j \zeta_j(t) + \zeta_j x(t)$$

(3)

$$p_j(t+\Delta t) = \gamma_j p_j(t) + z_j(t)$$

(4)

where δt(=0.01 s) is the simulation time step and $\gamma_j = e^{-\tau_j \Delta t}$ is the $j-th$ basis decay factor, derived from a relaxation time constant τj. ζj is a scaling factor that equalizes the power of all bases. At this point, an expansion of the original signal $x(t)$ into a series of bases or transients with different temporal profiles is obtained. The second processing step consists in mixing those bases according to a weight vector $w(t)$ to generate an output signal ($ff(t)$):

$$ff(t) = w(t)_T p(t)$$

(5)

where $p(t) = [p_1(t), ..., p_N(t)]_T$ is the vector of the bases. The weight vector is adaptively set by means of an Least Mean Squares (LMS) or Widrow-Hoff update rule [Stearns, 1985] extended with an eligibility trace:

$$w(t+\Delta t) = w(t) + \beta\epsilon(t)p(t-\delta_x)$$

(6)

where, $\epsilon(t)$ is an appropriated error signal that is used to update the weights. The eligibility trace is implicit in the use of a delayed copy of the bases activity $p(t-d_x)$ for the update, with x indexing the type of stimulus processed: proprioceptive (p) or distal (d). In short, to update the weights the current error is associated with an activity on the basis signals δ_x seconds ago. With that, we assume that activity at time $t - \delta_x$ is the one that should have been used to trigger a reaction with sufficient anticipation to cancel the current error at time t. In general, we set both δ_d and δ_p greater than the error feedback delay (δ_s), implying that the extent of the anticipation goes beyond the transport (or error feedback) delay.

Configuration of the FEL and HSPC architectures

Both control architectures include the feedback controller and two feed-forward modules (distal and proximal) wired according to the heuristic of either predicting motor commands from sensory signals (FEL architecture), or predicting sensory signals from sensory signals (HSPC architecture).

In FEL, feed-forward modules act upon the plant and are supervised by the feedback reaction to the error in posture (fig. 3.2-B). In particular, the proximal module issues a feed-forward action in response to the impact learned by shifting the reactive action earlier in time, while the distal module similarly acquires a response that is triggered by the distal stimulus, and thus can precede the impact itself.

Let $ff_p(t)$ and $ff_d(t)$ be the outputs of the proximal and distal feed-forward modules; $x_p(t)$ and $x_d(t)$, their respective input signals; and $\epsilon_p(t)$ and $\epsilon_d(t)$, their respective teaching signals. The structure of the FEL architecture is determined by the following equations:

$$x_d(t) = i_d(t)$$

(7)

$$x_p(t) = i_p(t)$$

(8) and

$$\epsilon_d(t) = \epsilon_p(t) = \tau_{fb}$$

(9)

where Embedded Image and Embedded Image represent the cue (distal) and impact (proximal) signals, respectively, and τ_{fb} is the output of the feedback controller. As a final step, the output of all modules are added up to generate the control signal ($\tau_{fel}(t)$):

$$\tau_{fel}(t) = \tau_{fb}(t) + ff_p(t) + ff_d(t)$$

In HSPC, upstream modules drive and learn from the input of downstream modules (fig. 3.2-C). That is, the proximal module learns *counterfactual* errors [Herreros et al., 2016] contingent to the impact so that the feedback controller reacts to the expected error before the actual one occurs. While the distal module learns to predict the collision signal contingent to the cue and triggers the proximal module ahead of the impact. Note that, by necessity, the HSPC architecture includes an internal comparator that computes the prediction errors associated with the collision signal.

In keeping with the above notational conventions, the equations determining the distal feed-forward module inputs and error signals in HSPC are:

$$x_d(t) = i_d(t)$$

(11) and

$$\epsilon_d(t) = i_p(t) - ff_d(t - \delta_d)$$

(12)

Note that the error signal that controls learning in the distal feed-forward module is a prediction error, coding the difference between a past prediction, $ff_d(t - \delta_d)$, and the actual stimulus, $i_p(t)$, where δ_d is the anticipatory delay of the distal module. The proximal feed-forward module is integrated within the control architecture as follows:

$$x_p(t) = \epsilon_d(t) + ff_d(t)$$

and

$$\epsilon_p(t) = -\theta(t - \delta_s)$$

In brief, the Sensory Prediction Error (SPE), $\epsilon_d(t)$, and the prediction signal, $ff_d(t)$, related to the collision drive the proximal module, which is supervised by the error in angle (measured with a delay of δ_s seconds).

In the last stage, the output of the proximal feed-forward module is added to the error in velocity driving the feedback controller. We formulate that operation by introducing $\epsilon_\theta(t) = -\theta(t-\delta_s)+ff_p(t)$ and then rewriting the first equation of the feedback controller:

$$\tau_{fb} = k_p\epsilon_\theta(t) + k_d\dot{(\epsilon_\theta)}(t)$$

(15)

Finally, the motor control signal generated by the HSPC architecture is simply the output of the feedback controller, $\tau_{hspc}(t) = \tau fb(t)$

3.3 Results

Below, we report on the performance of both the HSPC and FEL control schemes for three experimental conditions: standard acquisition trials, robustness (catch) trials in which the disturbance is cued but not delivered, and generalization trials in which we provide both cued and non-cued trials, and change the weight of the agent during training.

3.3.1 Acquisition

We start by analyzing the performance of the two adaptive control architectures in the acquisition of an APA trained in a trial-by-trial manner. We use a simulated self-balancing system that at each trial receives an impact, preceded by a distal cue by a fixed interval of 400 ms, and resulting in a disturbance force (100 N during 300 ms). The force, applied to the pendulum, produces an angular displacement that, in the naive system, is uniquely counteracted by the reactive controller introducing oscillations in the angular position (fig. 3.3-A, gray line). After learning, acquired motor responses evoked by the two predictive stimuli (cue and collision) substantially reduce the angular error (fig. 3.3-A, red and cyan). Note that despite implementing different adaptive strategies, we could configure both architectures to exhibit similar learning curves (fig. 3.3-B).

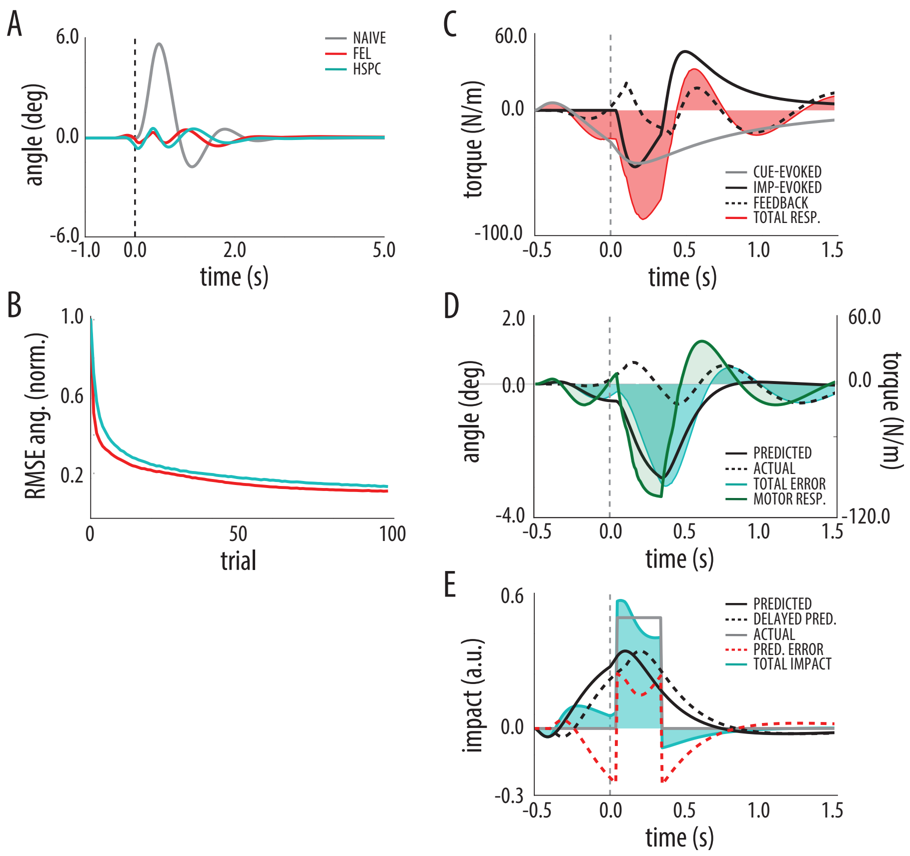


Figure 3.3: **Acquisition of adaptive postural adjustment. A.** Mean angular position during the disturbance rejection task for feedback-control condition (gray - 10 trials), trained FEL architecture (red - trials 90-100) and trained HSPC architecture (cyan - trials 90-100). Disturbance is delivered at t=0 (dashed line). **B.** Root mean square error (RMSE) in angular position over trials for FEL (red) and HSPC (cyan) architectures normalized by the maximum error in the naive system (feedback-control only). **C.** Decomposition of the motor response driving behavior in FEL: cue-evoked (gray), impact-evoked (solid black) and reactive (dashed black) responses are integrated in a total motor command (red shaded area). **D.** Decomposition of the angular error driving the behavior of the feedback controller:

the total error (shaded cyan area), obtained by summing the counterfactual (solid black) and the current error (dashed black), is converted into a motor response (green shaded area) by the feedback controller. Error and motor response refer to different scales. **E.** Decomposition of the impact sensory signal entering the feed-forward compensatory layer of HSPC: the total impact signal (shaded cyan area) is obtained as the sum of the predicted impact signal (black solid) and the prediction error (red dashed), which is computed by subtracting the delayed prediction (black dashed) from the actual impact signal (gray solid).

After learning, in FEL the reactive controller is only marginally engaged as the errors in behavior that drove it initially are almost canceled (electronic supplementary material, fig. S1a). Note that in this architecture, only the cue-evoked command contributes to preparatory behavior (before the collision) but both cue- and collision-evoked commands contribute to the fast feed-forward compensation that takes place after the collision (fig. 3.3-C).

Conversely, in HSPC the proximal adaptive module that associates the collision signal with inertial errors steers the feedback controller both during anticipation and fast compensation (fig. 3.3-D). Still, after learning, the proximal module is fed with a mixture of actual and anticipated collision signals, where the former is sensed and the latter provided by the distal module. Importantly, the distal module predicts the collision signal from the cue and issues an anticipated impact signal preceding the actual impact by 100 ms (the extent of the anticipation, δd, is a design parameter - see Methods) (fig. 3.3-E). Hence, the anticipatory part of the response, despite being evoked only by the cue stimulus, results from a cascade of predictions that involves both adaptive feed-forward modules and the feedback controller.

In sum, despite the marked differences in the processing, both architectures converged to similar motor commands and behavior, indicating that both motor anticipation- (FEL) and sensory prediction-based (HSPC) strategies can be equally successful in acquiring APAs.

3.3.2 Robustness

Next, we assess the reaction of both architectures to violations in the sequence of predicted events that was learned during training. To that end, after 100 acquisition trials, we run 50 trials within which we randomly intersperse 10% catch trials in which we present the cue but omit the disturbance. During catch trials, the agent initiates an anticipatory motor response that later, due to the lack of disturbance, results in a performance error [Lang and Bastian, 1999, Albert and Shadmehr, 2016, Witney et al., 1999]. Here, we use such errors to quantify how responsive FEL and HSPC are in recovering from erroneous predictions [Witney et al., 1999].

Prior to the expected impact time, both architectures introduce a slight anticipatory angular error (fig. 3.4-A) by issuing the preparatory part of the response (fig. 3.4-B) . However, once the impact fails to occur, HSPC promptly corrects the initial error while in FEL the error keeps increasing. In terms of performance, the error in a catch trial incurred by HSPC (median of the RMSE) is approximately half of the error introduced by FEL (0.3 versus 0.6 in normalized RMSE; fig. 3.4-C). The errors seen in catch trials are the same ones observed at the onset of extinction training. Both architectures greatly suppress these errors (also called after-effects) after 50 extinction trials (fig. 3.4-C).

The reasons behind the difference in performance in catch trials are the following: FEL reacts to the absence of the impact by omitting the collision-evoked command, but maintains the whole cue-evoked command even after the lack of the expected collision has shown it to be unnecessary. By contrast, HSPC rapidly aborts the (feed-forward) action once the proximal module receives the SPE triggered by the missed collision (fig. 3.4-D).

In summary, the HSPC architecture outperforms the FEL in that, due to the computation of sensory prediction errors, it can react on-line to violations in the course of expected events (i.e. to SPEs).

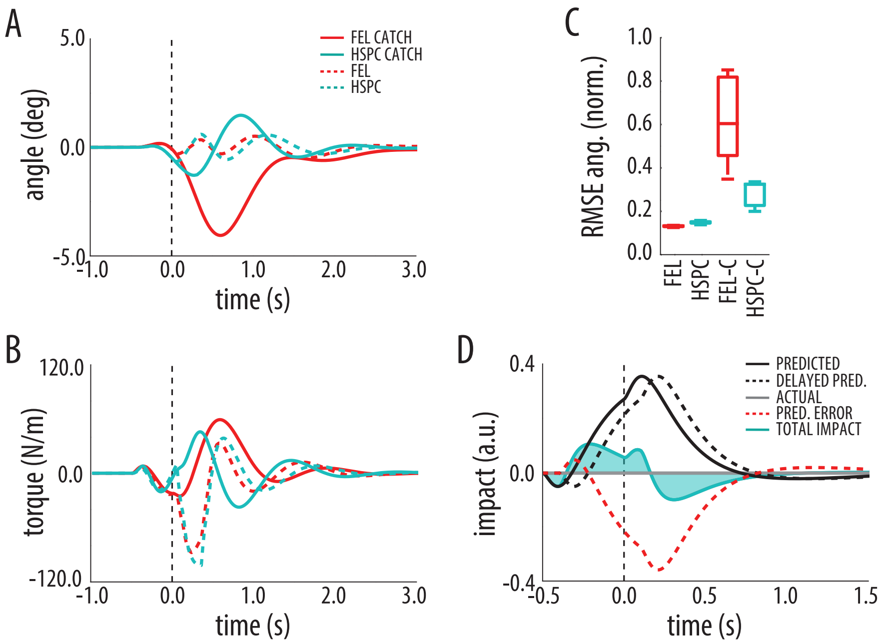


Figure 3.4: **Robustness of the FEL and HSPC architectures. A.** Mean angular position of FEL (red) and HSPC (cyan) during catch (N=10 - solid) and regular perturbed trials (N=10 - dashed). **B.** Mean motor response during catch and regular perturbed trials. Color-code as in A. **C.** Root mean square error (RMSE) in angular position during regular trained perturbed trials (N=40) and catch trials (N=10). **D.** On-line prediction error correction in HSPC: the prediction error (red dashed), obtained as the difference between the delayed prediction (black dashed) and the actual impact signal (gray solid) is subtracted from the erroneously anticipated impact signal (black solid) and generates a total response (cyan area).

3.3.3 Generalization

In a final set of simulations, we test how both architectures respond to changes in the plant dynamics and task contingencies. We run an additional set of 60 trials after acquisition. During the first 10 extra trials, we measure the performance of the feed-forward compensatory layer in isolation, omitting the cue. At trial 11, the plant is made heavier (+10% - light-to-heavy condition; note a similar manipulation in behavioral postural control studies [Li and Aruin, 2007]) and the agent receives additional non-cued collisions (40 trials). Afterwards, we reintroduce the cue for 10 more trials. In a separate set of simulation, we train initially the heavier agent and afterwards remove the weight (-10%-heavy-to-light condition).

In FEL, any change in the task decreases performance (removing or reinstating the cue), irrespective of whether the plant has increased or decreased its weight (fig. 3.5-A,C,E,F). In HSPC, the performance deteriorates, albeit to a lesser extent, after removing the cue. However, once the cue is reintroduced after having retrained the compensatory module, we observe a gain in performance in both cases, greater when transitioning to the lighter plant (fig. 3.5-B,D,E,F).

The difference in performance stems from the different ways in which both architectures combine the two stimuli. FEL deals with the cue and impact as independent stimuli. Initially, both contribute to the response, but once the cue is removed a part of the response is removed as well, damaging performance (electronic supplementary material, fig. S3a). Further training makes FEL able to trigger appropriate compensatory responses just with the proximal stimulus, but then, reinstating the cue superposes a motor command partly redundant, damaging performance again (fig. 3.3-C). Notably, if one would consider that cue and impact form a compound stimulus in regular trials, one could explain the interference between the cue and the impact stimuli with the Rescorla-Wagner model [Rescorla et al., 1972]. On the contrary, in HSPC the distal module learns to predict the impact from the cue, and uses that prediction to trigger (a part of) the compensatory action in anticipation (electronic supplementary material, fig. S3b). That implies that even after changing the properties of the plant, anticipating an appropriate compensatory action can result in an improvement in performance (fig. 3.3-D).

In summary, in face of perturbations to the plant dynamics or changes in the task contingencies, a control strategy learning a cascade of sensory predictions allows for better generalization than one that treats the different stimuli independently.

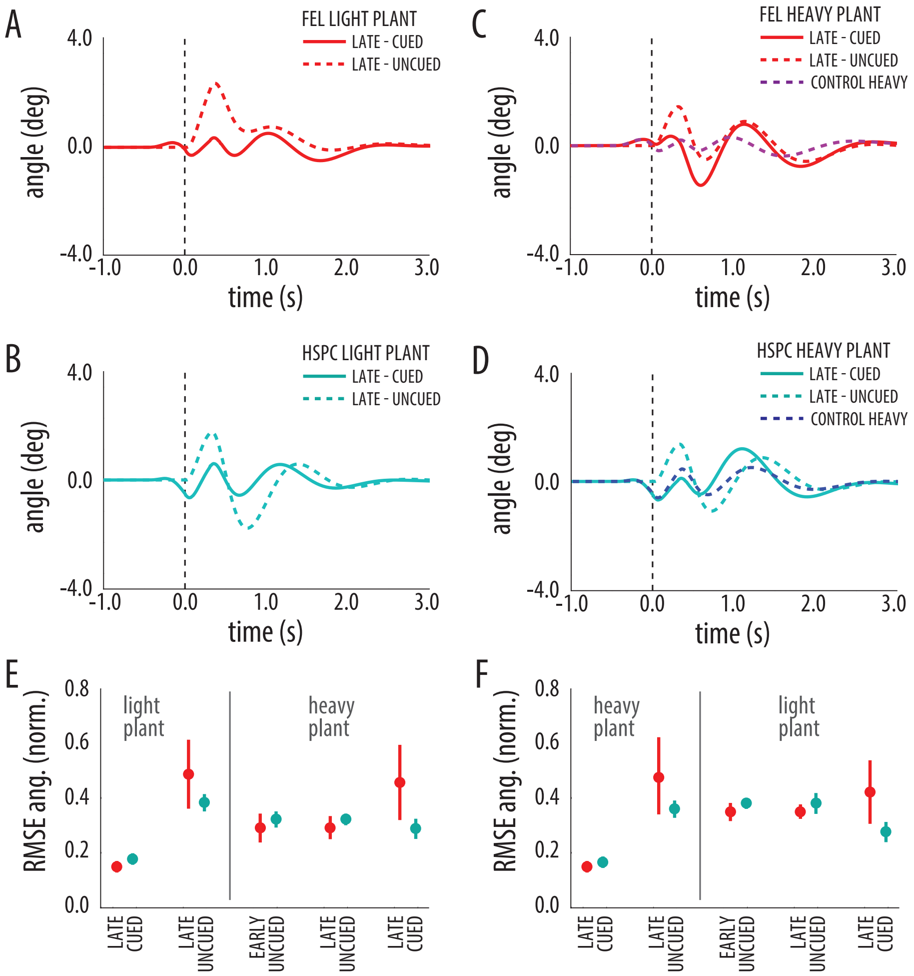


Figure 3.5: **Generalization. A.** FEL: Mean angular position before plant perturbation (N=10 - light plant) with (solid) and without the cue (dashed) **B.** HSPC: Same description as A. **C.** FEL: Mean angular position after plant perturbation (heavy plant - N=10) without (dashed) and with the cue (solid) and after regular training with heavy plant (solid magenta). **D.** HSPC Same description as in C. **E.** Root mean square error (RMSE)

in angular position during light-to-heavy generalization phases for FEL (red) and HSPC (cyan). ”Light plant” denotes the phase before plant perturbation. ”Heavy plant” denotes the phase after plant perturbation. **F.** Same description as E for heavy-to-light generalization.

3.4 Discussion

Even though it is clearly established that skilled motor behavior relies on internal models, their nature is still under debate. The two prevailing views are that internal models can be either inverse models, mapping the desired sensory consequences into their required motor commands, or forward models, mapping motor commands into their predicted consequences. Here, we have challenged this dichotomy and advanced an alternative proposal (HSPC) that reformulates anticipatory motor control as a sensory-sensory learning problem. On this view, the predicted consequences of responses to (distal or proprioceptive) cues prescribe action or motor commands (that are mediated - or realized - by reflexes). This simplification and generalization of the ’standard model’ appeal to active inference, with an emphasis on estimating and predicting states of the world and the self. In order to test this hypothesis, we designed two control architectures that adopted either a motor anticipation- or a sensory prediction-based approach. We based the motor-anticipation architecture on the well-established FEL model [Kawato et al., 1987, Kawato and Gomi, 1992, Gomi and Kawato, 1993] whereas HSPC provided the sensory prediction-based architecture.

We compared both architectures in a simulated APA task [Lacquaniti and Maioli, 1989, Lang and Bastian, 1999, Santos et al., 2010b]. Despite differences in the processing, both architectures acquired an APA equally well (fig. 3.3-A,B). However, as soon as we extended the basic APA protocol with either the introduction of catch trials or by perturbing the plant, the sensory prediction strategy outperformed motor anticipation. Below, we will argue that the reasons for that superior performance are grounded in two specific consequences of the sensory prediction strategy:

its reliance on SPEs, and second, that HSPC affords a hierarchical processing architecture that encapsulates learning at different levels. In other words, in line with active inference, placing a hierarchical model on top of reflexive sensorimotor control equips behavior with a context-sensitivity and intentional aspects that are precluded in ’standard’ formulations.

3.4.1 Origin of the robustness and generalization capabilities in HSPC

The hierarchical structure of the HSPC explains its superior generalization ability. The FEL architecture has a flat structure as far as controlling behavior is concerned: all modules send motor commands in parallel to the plant. This means that after a perturbation of the plant, the output of all modules has to be retrained to the new plant dynamics. In HSPC, its hierarchical structure entails that all modules are only concerned with driving and learning from the module immediately below in the hierarchy. Hence, HSPC solves the control problem by partitioning it into two smaller sub-problems: predicting the collision from the cue and predicting the postural errors from the collision. As a consequence, changing the mass of the agent only changes the sensory consequences of the collision, hence, once a new feed-forward reaction to the collision is acquired, a gain in performance can still be obtained by correctly anticipating the collision (thereby, bringing the trained reaction forward in time).

On the other hand, SPEs enable the fast reaction to erroneous predictions. As FEL only learns to react to stimuli, but not to predict them, it cannot (at least naturally) incorporate SPEs. On the contrary, HSPC relies on SPEs both for improving prediction accuracy and to preclude reaction to predicted stimuli at the time of their actual occurrence [Miall et al., 1993]. That is, SPEs are intrinsic to the design principle behind HSPC. In catch trials, as no collision occurs, the prediction of the distal module fails, generating a negative SPE that interrupts the ongoing response of the proximal module initiated by the distal module, thereby enabling a fast recovery (in addition to readjustment - learning - as the absence of the collision may imply a lasting change in task contingencies).

3.4.2 Environmental forward models and inverted sensory-sensory forward models

The distal module in HSPC is a forward model of the environment that solves the problem of predicting one stimulus (a collision) given another stimulus (a cue); that is, a task contingency. In general, forward models of the environment have been acknowledged [Miall and Wolpert, 1996], but usually not considered specifically in the context of physiological motor control except, recently, within the domain of active inference [Serrien and Wiesendanger, 1999b, Friston, 2011a, Friston et al., 2010, Herreros and Verschure, 2015]. However, the forward model in HSPC is not generically predicting one stimulus from another; it is anticipating a stimulus with the objective of driving a behavioral response that minimizes a defined error. For that, it must take into account not only sensorimotor latencies but also the dynamics of the plant (e.g. musculo-skeletal system). Hence, the environmental forward model in HSPC affords action-aware sensory predictions in that they are made having knowledge about the dynamics of the action that they will drive. By contrast, standard forward models do not require knowledge of the dynamics of the feedback action itself, as they only need to be tuned to the afferent and efferent delays [Miall et al., 1993].

On the other hand, the internal model dealing with the collision signal acts as an inverse model. Even though it is supervised by a postural error signal, its goal is not to learn to predict postural errors, but to steer its downstream feedback controller to avoid these errors. We have earlier called this approach counterfactual predictive control (CFPC) [Herreros et al., 2016]. The goal of CFPC is acquiring counterfactual error signals that, even though they do not code any forthcoming errors derived from the interaction with the physical world, they are processed by a feedback controller as if they were real errors. In practice, this leads the adaptive model within the HSPC architecture to acquire an inverse model of the closed-loop system that reflects jointly the dynamics of the plant and the controller [Herreros et al., 2016]. That is, a model is said to be inverse because it reverses a causal relationship: from the desired effects (i.e.

avoiding errors in performance) to inferring the right causes (i.e. the motor commands that will avoid those errors). The module processing proximal events within HSPC shares the same goal as the standard inverse model just mentioned. The only difference is that it outputs a predicted sensory signal that signifies an error rather than a motor command. This signal must be considered counterfactual. This demonstrates how a learning process that depends on sensory errors (in contrast to motor errors) is not automatically building a forward model (for another example, see [Tseng et al., 2007]).

3.4.3 Related research in experimental psychology and predictions of the HSPC hypothesis

Experimental APA protocols include standing human participants receiving the impact of an object attached to a pendulum [Santos et al., 2010b, Mohapatra et al., 2012]. As expected, those experiments show that faced with the incoming pendulum, participants rely on distal sensing (vision) to issue the anticipatory responses [Santos et al., 2010b, Mohapatra et al., 2012], that is: no anticipatory responses were observed when participants closed their eyes. Regarding the interplay between proprioceptive and vestibular information, separate studies in compensatory postural control have shown that humans with compromised proprioception display compensatory responses delayed with respect to healthy controls [Allum et al., 1998] as well as animals with pyridoxine-induced loss of peripheral sensory efferents have delayed compensatory responses and increased postural sway [Stapley et al., 2002]. This suggests that, despite some simplifications, the design of the task and the adaptive interplay between sensory modalities and responses in our simulated APA task is in close agreement with well-studied properties of biological control. We note, however, that in humans and animals, anticipatory and compensatory strategies often act synergistically across different sets of muscle synergies, reflecting different demands (i.e. upper extremities respond with a higher degree of anticipation compared to lower ones) [Iodice et al., 2015]. However, those findings do not discriminate between the sensory prediction and motor-anticipation hypotheses. An exception comes from experiments showing that altered proprioceptive information at the level of the Achilles tendon delays anticipatory postural responses [Mohapatra et al., 2012]. Note that FEL would predict that decreasing the information in the proprioceptive channel would have no effect in the preparatory actions, which are motor commands triggered by the visual stimulus. However, in the HSPC hypothesis, anticipatory actions are elicited by generating proprioceptive predictions. Hence, one could expect that a manipulation that alters the processing of real proprioceptive information would also affect the mapping of predicted proprioception into action.

HSPC further predicts that in catch trials, subjects will correct erroneous anticipatory actions with a latency equal to the time needed to detect SPEs. By contrast, as FEL makes no use of SPEs, it has no mechanism that could detect and process such a sharp change in behavior at the expected time of the disturbance. Note that errors observed in catch trials, or after-effects, which provide a means to quantify learned motor responses, are a hallmark of adaptive motor behavior. Hence, as HSPC greatly diminishes those after-effects, it may seem that we are advancing a control scheme whose performance is non-biological. However, this is not the case, first, because HSPC reduces after-effects but it does not suppress them but rather they are subject to extinction, or washout (fig. 3.4-A,C). Second, HSPC curtails erroneous feed-forward responses as soon as SPEs can be detected. Experimentally, the fast correction in catch trials that we demonstrate with HSPC has also been observed using a grip-force modulation paradigm where participants learned to anticipate an artificially delayed (but self-generated) disturbance [Witney et al., 1999]. However, to the best of our knowledge, this kind of catch trial has not been studied in the context of the anticipatory control of balance. For an APA task as the one we have modelled, providing catch trials will likely require a virtual reality setup allowing to decouple the distal and proximal cues; that is, showing a virtual looming object that in paired trials coincides with an actual object hitting the participant but that in catch trials does not.

In addition, generalization of adaptive motor responses has been found in limb [26,54] and postural control [Ahmed and Wolpert, 2009]. Subjects trained to catch a ball with one arm perform equally good when they switch arm [Morton et al., 2001], a result that cannot be explained in terms of inverse models (by definition, effector specific). Moreover, subjects that learned to counter a force-field perturbation in a sitting position correctly anticipated the postural disturbances that compensating for the force field would introduce in an upright posture [Ahmed and Wolpert, 2009]. This result argues in favor of an architecture composed of a forward internal representation of the dynamics of the environment coupled with an internal model of the postural dynamics, where the former is effector independent and the latter is already fine-tuned by experience; a proposal consistent with the hierarchical structure of HSPC.

Put together, these three sources of evidence (generalization of acquired responses across limbs and postures, rapid reversal of the erroneous response in catch trials and anticipatory responses affected by altered proprioception) support a hierarchical control architecture that acquires forward models of the environment, exploits SPEs and shows a dependency between anticipatory and compensatory responses. All these features are embodied in HSPC but are difficult to reconcile with an inverse model-based architecture such as FEL.

Finally, APAs are also observed in response to voluntary actions that trigger self-generated perturbations (e.g. extending an arm, loading a weight) [MacKinnon et al., 2007, Cordo and Nashner, 1982]. Even though we focused on externally generated perturbations, HSPC could account for self-initiated perturbations by replacing the distal sensory input with an internally generated signal encoding the initiation of a motor plan [Ruck et al., 2016], which would trigger a similar cascade of sensory predictions.

3.4.4 Implications for cerebellar physiology

HSPC advances a hypothesis of cerebellar function in the domain of anticipatory control. It has its origins in a model of the cerebellum [Herreros et al., 2016, Herreros and Verschure, 2013a] as is the case for FEL [16]. In both architectures, adaptive modules are implemented as adaptive filters, a widely used computational model of cerebellar function [Dean et al., 2010a, Fujita, 1982]. Moreover, here we have demonstrated HSPC in a task that depends on the cerebellum [Kolb et al., 2004, Timmann and Horak, 2001a, Diedrichsen et al., 2005]. A distinctive trait of our implementation of the cerebellar algorithm is the use of a delayed eligibility trace (Methods - equation (2.1)) [Herreros and Verschure, 2013a]. Taking into account that in the cerebellum contextual information reaches Purkinje cells through the parallel fibers whereas specific error signals arrive via the climbing fibers, in terms of cerebellar physiology, the eligibility trace mechanism predicts a plasticity rule in the synapses between parallel fibers to Purkinje cells that modifies synaptic weights whenever activity in the parallel fibers precedes climbing fiber input by a certain time interval. Both in HSPC and FEL, we set that interval according to the behavioral constraints of the agent/task [Herreros et al., 2016], a requirement that seems to apply also in the cerebellum, where the timing of the plasticity rule of cerebellar Purkinje cells is matched to behavioral function [Suvrathan et al., 2016].

From a system level perspective, our proposal emphasizes the computations that could be achieved by organizing cerebellar modules in a hierarchical fashion. At the level of anatomy, such a functional hierarchy would require cerebellar microcircuits to be serially connected. That is, the output from one microcircuit could provide an input to the next one (or ones) in the hierarchy. This could be realized as non-reciprocal nucleo-cortical connections by which a particular area of the cerebellar nucleus could feed a cerebellar cortical microzone projecting to a separate region of the cerebellar nucleus. This arrangement is in agreement with the descriptions of the organization of the nucleo-cortical projections between the nucleus interpositus posterior (NIP) and the nucleus interpositus anterior (NIA) already present in the literature [Apps and Garwicz, 2005]. Indeed, it has been shown that a proportion of nucleo-cortical projections originating in NIP target NIA, whereas the opposite is not the case. This would imply that activity on NIP could modulate, after one step of cerebellar cortical processing, activity in the NIA. On a more speculative note, from the perspective of HSPC, we would expect NIA to be more directly involved in motor control tasks (i.e. targeting motor nuclei) [Monzée and Smith, 2004] whereas NIP would be more linked to sensory processing areas. Indeed, tracing studies showed that NIP sends its outputs to the ventro-lateral and ventro-posterior nuclei of the thalamus [Angaut, 1970], main relays to somatosensory cortical areas crucial for the computation of SPEs [Mathis et al., 2017] and to frontal cognitive areas [Lu et al., 2012, Ramnani, 2006a].

3.4.5 Summary

We have shown how a hierarchical control architecture based on sensory predictions enables the acquisition of responsive and generalizable APAs better than one based on the traditional view building on sensory-motor associations. In doing so, we went beyond the standard inverse-forward model dichotomy by showing how (the inversion of) forward models that acquire sensory-sensory associations can contribute to motor behavior with, what we have called, action-aware sensory predictions. Our results provide a validation of key principles behind the active inference framework of motor behavior and their realization in the DAC theory. In future work, we shall study how this anatomically constrained theory of anticipatory motor control could be extended to address the questions of optimality that arise when one takes effort-error trade-offs or the modulation of task-irrelevant versus task-relevant variability into account [Todorov and Jordan, 2002]. At this point, however, we expect the HSPC architecture to allow for the advancement of our understanding of the mechanisms underlying physiological anticipatory motor control, which we propose can now be treated in a framework related to active inference, while also contributing to the development of robust control architectures for artificial systems.

Part III

Chapter 4

DELIBERATE CONTROL OF ACTION IN HUMAN MEDIAL FRONTAL CORTEX

In this chapter we extend the notion of prediction violation from the previous chapters by introducing the notion of goal, that is, we focus on the type of violation that demands not just to halt behavior, but to switch to a different behavioral objective. The aim is to elucidate what cortical area and what features of the neural activity encode the control signal promoting the switch from automatic to deliberate control within a human intracranial EEG setup. The task, based on the standard serial reaction time task, requires the subjects to perform an habitual (feed-forward) sequence of button presses, occasionally interrupted by the appearance of a cue that requires them to press an alternative, un-cued, key. We find that medial frontal areas (SMAs) are involved in switching from automatic to deliberate control. In addition, we suggest that phase alignment of the oscillatory frequencies in the theta band, the synchronization with temporal cortices and the coupling with higher frequencies (gamma) contributes to explain behavioral performance. Altogether, these results may suggest that frontal theta oscillations could implement an actual control signal through which executive and memory areas retrieve and execute goal-oriented actions.

This chapter is based on:

Maffei, G., Puigbo, J. Y., Santos-Pata, D., Zucca, R., Principe, A., Roccamora, R., Conesa, G., and Verschure, P. (2018). Theta phase mediates deliberate action switch in human SMAs. (in preparation)

4.1 Introduction

The ability to deliberately overwrite ongoing automatic actions as a response to contextual changes is a necessary feature of animal behavior, with important implications for survival. The Supplementary Motor Complex (SMC), in the medial frontal cortex, is thought to mediate the switch from automatic to deliberate control when a detected conflict between current and expected contingencies requires a change in the ongoing motor plans [Rushworth and Hadland, 2002] **[Hikosaka 2004]**.

The hypothesis that SMC could work as a neural controller that orchestrates automatic and deliberate processes [Hikosaka and Isoda, 2010a] requires this area to accomplish the following functions: (a) detect a sudden contextual change, (b) inhibit ongoing motor routines, and (c) facilitate the execution of alternative "deliberate" actions. As an additional requirement, we argue that the SMC should also be able to access internal representations (i.e. memory) to inform behavioral choices in the case when environmental cues do not indicate what alternative action should be taken.

Frontal oscillations in the theta range (4-8 Hz) have been previously reported to mediate cognitive control [Voytek et al., 2015] [Helfrich and Knight, 2016] suggesting that dynamics within this restricted set of frequencies may reflect a control signal that meets the requirements postulated by theoretical models. In particular, it has been shown that theta oscillatory phase mediates the communication between parietal and medial frontal cortical areas during actions switch in primates [Phillips et al., 2014], a mechanism potentially necessary to detect sudden contextual changes. In addition, a number of studies have implicated the power of theta oscillations in inhibitory control. For example, multiple human EEG studies have shown a positive relationship between theta power and response time during high-conflict trials, suggesting an implication of this frequency band (often termed ’frontal theta’) in the inhibiting rapid (automatic) responses and acting as a break for the motor system both prospectively and retrospectively (i.e. post-error slowing) [Cavanagh and Frank, 2014].

Interestingly, however, a neural signature in SMC, which reflects a control signal that facilitates action during deliberate action switch, has remained elusive. This is a pressing issue for the study of the SMC since it renders current interpretations of its functions, especially in humans, theoretically appealing but experimentally incomplete. A further experimental limitation that restricts the interpretation of the role of SMC is the poor characterization of the functional network that underlies deliberate actions switch [Nachev et al., 2008]. For example, it remains unclear what are the areas involved in this behavior during situations where the alternative objective is not explicitly cued and it requires to be internally retrieved.

In this study, we aim at identifying the neural features that reflect a facilitatory control signal of action switching in the human SMA and to characterize the functional network involved in this behavior. Three subjects implanted with intra-cranial electrodes in the medial frontal cortex underwent a variation of the Serial Reaction Time Task (SRTT) [Nissen and Bullemer, 1987], a paradigm that requires the execution of a sequence of repetitive visually-guided key-presses, which is progressively automated. In a small subset of trials, the automatic sequence was unpredictably interrupted by the appearance of a cue (switch trials). During switch trials, subjects were required to interrupt the ongoing action sequence and press an alternative un-cued key.

Consistently with previous reports [Isoda and Hikosaka, 2007,Hikosaka and Isoda, 2010b], we found that the SMC is involved in the control of action during switch trials but not during the automatic ones, with the time to peak of the evoked potential (ERP) being predictive of the response time. However, once temporal differences between ERPs were filtered out by aligning the trials to an endogenous (ERP peak), rather than an exogenous event (stimulus presentation) [Voytek et al., 2015], our analysis revealed no significant difference in neither peak amplitude nor oscillatory power between fast and slow trials. In contrast, we found a significantly higher phase alignment in the theta band (4-9 Hz) in fast trials compared to slow trials. Further, single trial coherence analysis [Hipp et al., 2011, Jarvis and Mitra, 2001] confirm this negative relationship and propose phase dynamics as a reliable predictor of prompt response time. In addition, we report a significant increase of theta-gamma phase-amplitude coupling associated to faster responses suggesting a modulatory effect of theta rhythms on local population activity. Finally, to characterize the functional network underlying deliberate action switch we computed inter-area phase coherence during switch trials. This analysis revealed a higher synchronization between medial frontal cortex and temporal lobe during fast trials compared to low trials and suggests a critical contribution of memory to the retrieval and execution of appropriate alternative responses.

Altogether, our results confirm the role of human SMAs in automatic to deliberate action switch and support the involvement of the oscillatory dynamics in cognitive control. Differently from previous results, however, here we argue in favor of a direct implication of phase dynamics of the theta band in facilitating the execution of deliberate actions, a novel observation which supports theoretical accounts of SMC functions. In summary, we provide a novel evidence for the role of oscillatory dynamics in action execution in the human brain. This outcome supports a theoretical model of SMA as a controller in charge for, not just inhibiting concurrent motor plans, but also facilitating deliberate execution and suggests a functional pathway underlying behavioral performance that involves executive and memory network in the control of goal-oriented action.

4.2 Results

4.2.1 Behavioral results

In order to explore the neural dynamics underlying behavioral performance in deliberate action switch, three human subjects implanted with intracranial electrodes (iEEG) in the supplementary motor areas (SMAs), underwent a variation of the serial reaction time task [Nissen and Bullemer, 1987] **(Lee 2004)** (fig. 4.1-A). The first phase of the task (trial 0-60) required the participants to learn to perform a repetitive sequence of key-presses (N=5) on a touch-screen keyboard by following the presentation of a visual cue (green) until they reached automaticity. Automaticity was defined as the decrease of inter-key-intervals (IKI) to asymptotic value, indicating that the subject effectively internalized the motor sequence and relied less on visual feedback [Nissen and Bullemer, 1987] (fig. 4.1-C). In addition we observed a significant decrease in IKI mean time (fig. 4.1-D)(t-test ind: S1, $t = 5.30, p < 10^{-05}$; S2, $t = 5.71, p < 10^{-05}$; S3, $t = 4.08$, $p < 10^{-03}$) and IKI variability (fig. 4.1-E) (t-test ind: S1, $t = 2.21$, $p = 0.04$; S2, $t = 2.45, p = 0.02$; S3, $t = 3.28, p = 0.004$) by the end of the learning phase, suggesting that the subjects performed movements in a more stereotyped (i.e. automatic) manner [Sakai et al., 2004]. In the second phase of the task (trial 60-565), subjects were required to perform the same sequence of visually guided key presses but they were unpredictably interrupted by an appearance of a switch cue (red) at pseudo-random intervals (7 ± 2). When the switch cue appeared, subjects had to interrupt the ongoing automatic motor sequence and press an uncued key (instruction given at the beginning of the experiment) (switch trials). Subjects were able to successfully interrupt the ongoing action sequence during most of the switch trials, exceeding the key-press after cue presentation only in few cases (S1, N=7; S2, N=11; S3, N=13). Nevertheless, we report a systematic increase in response time during switch key-presses compared to automatic ones (fig. 4.1-F) (t-test ind: S1, $t = -39.59$, $p < 10^{-120}$; S2, $t = -32.96$, $p < 10^{-106}$; S3, $t = -71.56$, $p < 10^{-256}$) and large variability in switch response times (calculated from switch cue

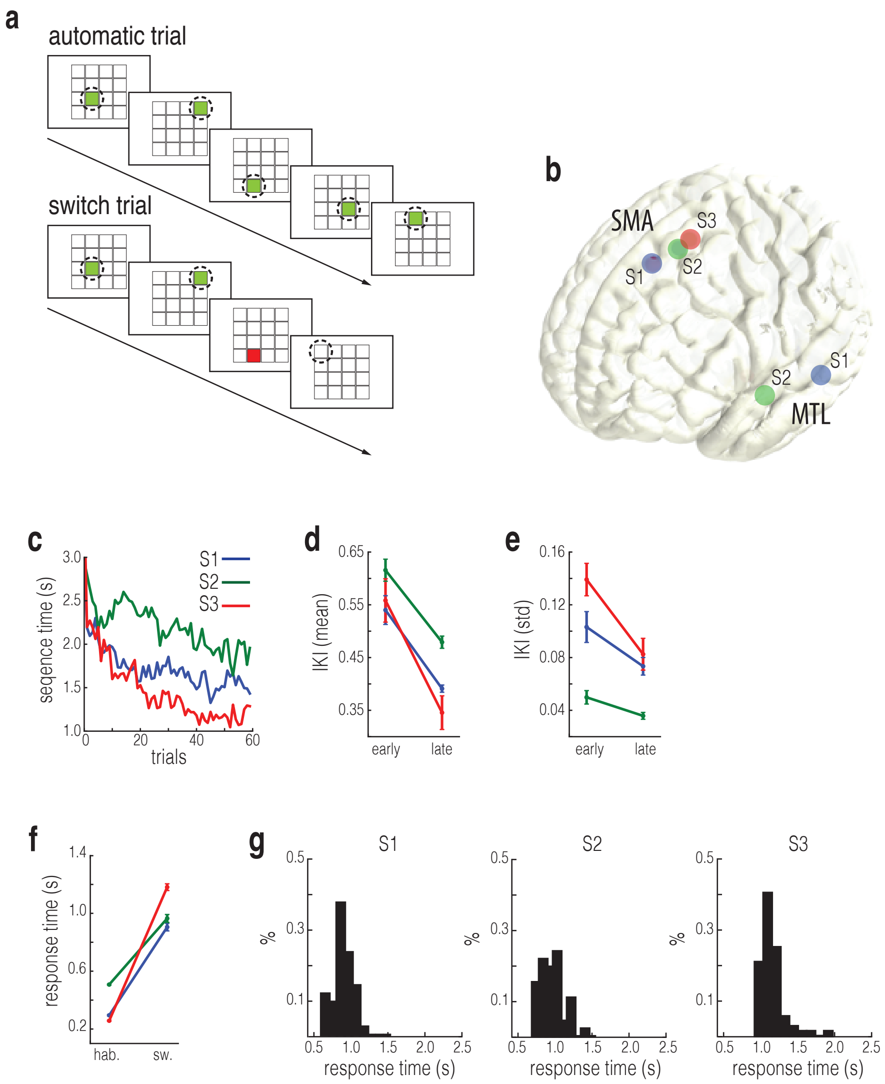


Figure 4.1: **Setup and behavioral results. a.** Serial reaction time task. During habitual trials subjects are required to perform a series of visually guided key presses (green key) following a pseudo-randomly generated sequence. During switch trials subjects are presented with a switch cue (red key) appearing at a random moment of the sequence requiring them

to interrupt the ongoing motor sequence and press an uncued key. **b.** Projection of relevant contact points locations over MNI atlas for each subject. **c.** Sequence time evolution of the performed sequence during training for each subject. **d.** Mean inter-key-interval (IKI) during early (trial 1-10) and late (trial 50-60) phases of the training for each subject. Errorbars indicate S.E.M.**e.** Inter-key-interval (IKI) standard deviation during early (trial 1-10) and late (trial 50-60) phases of the training for each subject. Errorbars indicate S.E.M. **f.** Response time of habitual and switch trials for each patient. Errorbars indicate S.E.M. **g.** Normalized distribution of response time during switch trials for each subject.

presentation to uncued key press), varying from approx 600 ms to 1300 ms (fig. 4.1-G). Such variability in performance cannot be fully explained either by a consistent learning effect across subjects or by the position of the switch cue within the sequence (showing a significant effect only in one subject), and it may, therefore, imply different characteristics of the neural mechanism underlying switch action control.

4.2.2 SMA is involved in switch but not automatic action

Based on previous literature [Rushworth and Hadland, 2002, Isoda and Tanji, 2004, Nachev et al., 2008], we hypothesized that frontal circuits and in particular SMAs could be involved in the switch from automatic to the deliberate control of action. We confirm this hypothesis by performing a time domain whole-brain classification on the available electrodes for each patient and show that medial-frontal cortex plays an active role in this task, being the features of the neural signal predictive of the type of trial (automatic vs switch) (fig. 4.2-A). Following this step, we restrict our analysis to the relevant contact points in the SMAs. Time domain analysis of the neural signal reveals a significant increase of LFP activity locked to the switch action during switch trials, but not during automatic trials (T-statistics cluster permutation ($N = 1000$) analysis: S1, $t = 34.09$, $p < 10^{-4}$; S2, $t = 21.20$, $p < 10^{-4}$; S3, $t = 29.69$, $p < 10^{-4}$) (fig. 4.2-B). This suggests that SMAs are not generally involved in the control of

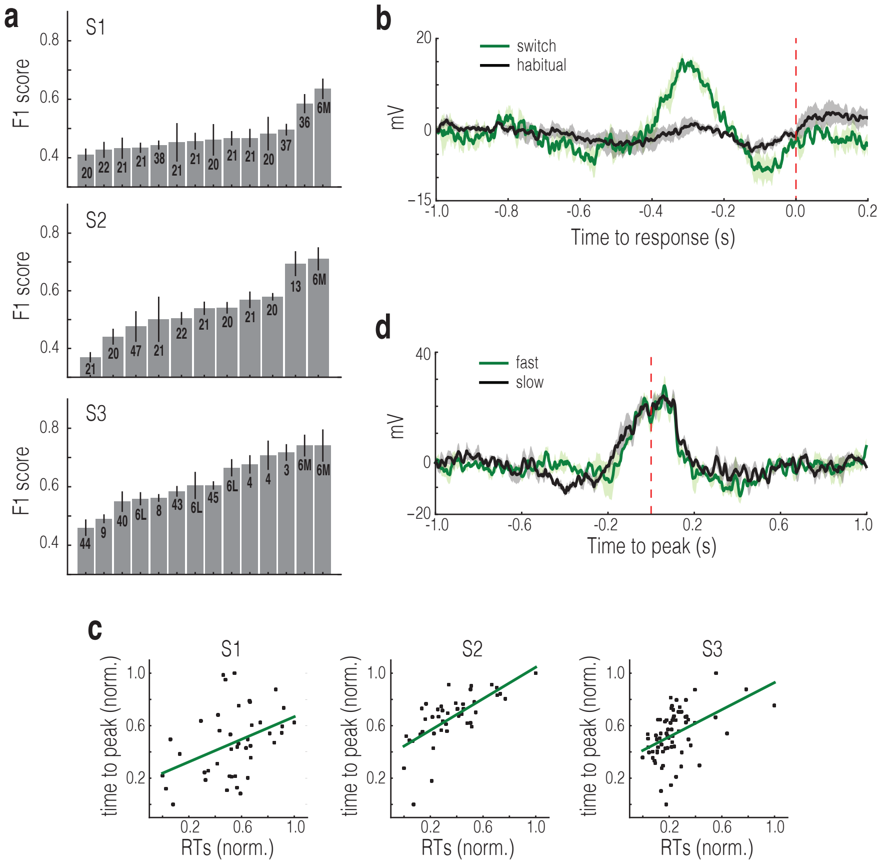


Figure 4.2: **Neural response in SMA.a.** Classifier prediction accuracy of switch trials across recorder areas. Labels indicate the mapping to Brodmann atlas for each contact point **b.** Neural response in SMA during switch (green) and habitual (black) actions aligned to key-press (red dashed) . Mean and SEM of subjects. **c.** Relationship between response time and erp time-to-peak (from cue presentation) during switch trials for each patient. Green line indicates the linear fit of the data. **d.** ERP in SMA during fast and slow response times (trial sorted by the median of the distribution) aligned to ERP peak (red dashed). Mean and SEM of subjects.

over-trained sequential motor responses, but might mediate the execution of those actions that require higher cognitive control [Rushworth and Hadland, 2002]. Further trial-by-trial analysis supports the behavioral relevance of the detected response, showing a positive correlation between the single trial time-to-peak from stimulus presentation and the relative response time (Person correlation coefficient: S1, $R = 0.406$, $p < 0.01$; S2, $R = 0.698$, $p < 10^{-8}$, S3, $R = 0.456$, $p < 10^{-5}$) (fig. 4.2-D). In contrast, peak amplitude showed no significant effect on response times (Person correlation coefficient: S1, $R = 0.118$, $p = 0.448$; S2, $R = 0.265$, $p = 0.074$, S3, $R = 0.161$, $p = 0.174$). Altogether, these results suggest the idea that SMAs are recruited principally for cognitive control of behavior and not while performing non-deliberative actions. The behavioral correlation with time to peak and not with peak amplitude further suggest an involvement of phase and not power in the spectral domain.

4.2.3 Theta phase aligns in faster actions

We hypothesized that oscillatory dynamics in the theta range could constitute a neural signature of cognitive control by facilitating action execution. Previous reports have suggested a strong implication of phase dynamics in cognitive control [Voytek et al., 2015, Phillips et al., 2014, Helfrich and Knight, 2016] suggesting the hypothesis that stereotypical phase profiles could underlie deliberate action modulation. To detect stereotypical phase patterns underlying trial differences in response time, we sort the switch trials for each subject into two classes of equal size (*FAST* and *SLOW*), by splitting the RT distributions by their median value. Further, to perform a comparison of trials with different temporal profiles we suppress inter-trial temporal differences by locking individual trials to an endogenous event (ERP peak) rather than to an exogenous one [Voytek et al., 2010]. In order to detect differences in phase alignment predictive of response time, we calculated the PLV for the separate classes of trials for each patient and computed the normalized phase coherence difference following the method described in [Maris et al., 2007] (see Methods).

Phase coherence analysis revealed greater phase alignment in fast trials compared to slow trials in the 5-7 Hz range for all subjects (Z-statistics cluster permutation ($N = 1000$) analysis: S1, $z > 2.58$, $p = 0.031$; S2, $z > 2.58$, $p = 0.038$; S3, $z > 2.58$, $p = 0.042$) (fig. 4.3-A,B). Note that a significant increase of phase coherence was detected earlier in S1 ($-0.4s$) and later for S2 ($-0.1s$) and S3 ($0.1s$), possibly reflecting individual differences in implant locations (more frontal for S1).

It has been argued that phase coherence may be induced by increases in the power of the oscillations, constituting, therefore, an evoked rather than a real phase alignment [Lopour et al., 2013]. We controlled for this possibility by performing a similar analysis on the normalized power spectrum (fig.4.3-C). Cluster-based permutation statistical test showed no significant differences in power between trial classes. If theta phase alignment modulated action execution, a significant relationship between RTs and PLV should be found over a continuous range of RTs. To answer this question, we computed single-trial phase coherence (STPC) by deducting the contribution of individual trials to the overall coherence value following a jack-knife procedure based on [Hipp et al., 2011, Jarvis and Mitra, 2001]. We restricted this analysis to the significant temporal windows in the 5-7 Hz range emerged from the class comparison, and related cluster average STPC of each trial with the respective RT. We find a significant relationship between STPC and RTs for all subjects (Person correlation coefficient: S1, $R = -0.399$, $p = 0.008$; S2, $R = -0.331$, $p = 0.027$, S3, $R = -0.340$, $p = 0.003$), where STPC decreases with the increase of response time (fig. 4.4-A). These results support a direct implication of theta phase in the modulation of action parameters under cognitive control demands.

[ht]

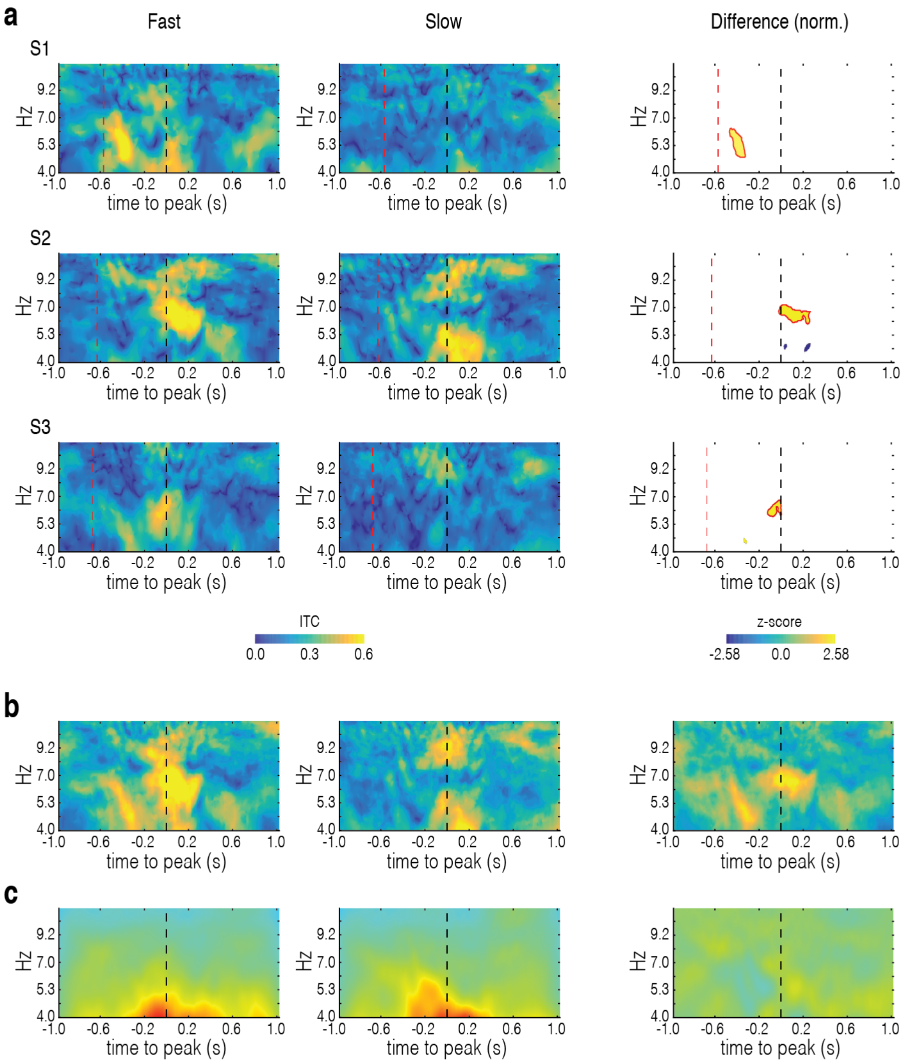


Figure 4.3: **Phase alignment differences in fast vs slow trials. A.** Phase alignment (inter-trial phase coherence) in fast (left) and slow (center) trials aligned to the peak of the ERP (black dashed) for each subject and normalized difference between trials (right). Red circle indicates $p < 0.05$. **B.** Mean phase alignment in fast (left) and slow (right) trials aligned to the

peak of the ERP (black dashed) and mean normalized difference (right). **C.** Mean normalized power in fast (left) and slow (right) trials aligned to the peak of the ERP (black dashed) and mean normalized difference (right).

4.2.4 Cross-frequency coupling predicts faster movements

We have suggested that theta phase alignment critically contributes to deliberate action execution, however, the physiological link needed to support that theta phase dynamics modulate behavioral performance via phase-dependent neural activity is still unclear. We sought to answer this question by determining the modulatory effect of theta phase on local high-frequency activity **(Bastos 2018)** with the hypothesis that higher modulation could support faster actions, an analysis for which a measure of cross-frequency Phase-Amplitude Coupling (PAC) is particularly suited. We restricted our analysis to the temporal window where a significant increase in phase alignment was detected, and, for each subject, we obtained one surrogate signal for fast and slow trials by concatenating the respective single trial windows, so to achieve the temporal resolution necessary for this type of analysis. Further, we computed PAC values (using the GLM approach [Penny et al., 2008]) between 5-7 Hz phase (modulatory frequency) and the amplitude of higher frequencies (10-100 Hz in steps of 2 Hz, modulated frequency), and obtained the difference between the two classes. This analysis highlighted an increased modulatory effect of the theta oscillatory phase on the amplitude of frequency bands in the gamma range (30-80 Hz), consistent across patients (Z-statistics permutation ($N = 1000$) analysis: S1, $z > 2.58$, $p = 0.027$; S2, $z > 2.58$, $p = 0.034$; S3, $z > 2.58$, $p = 0.029$) (fig. 4.4-B,C). In addition, a significant modulatory effect was found in the beta range (20 Hz) for one subject (S1, $z > 2.58$, $p = 0.019$).

This result supports a neurophysiological link between theta phase coherence and deliberate control of the action by means of modulation of high-frequency activity, often interpreted as a correlate of local population activity **(Bastos 2018)**.

4.2.5 SMA synchronizes with Medial Temporal lobe in fast trials

Whereas high-frequency amplitude may reflect local population activity, low-frequency oscillations could provide a mechanism for long-range synchronization between different neural populations [Hyafil et al., 2015, Fries, 2015]. The phase profile in the theta band emerged from the previous analysis could, therefore, represent a signature of a communication channel between the medial frontal cortex and the network of brain regions that are functionally involved in deliberate action switch. In particular, one of the distinctive features of our task is the absence of an explicit cueing on what action should be taken upon switch stimulus presentation and requires the subjects to retrieve this instruction from memory. It is, therefore, possible that the SMC could rely on memory areas in order to facilitate the retrieval of the appropriate action to be performed. We tested this hypothesis in two of the three patients (S1, S2) that had implants in the temporal lobe (TL), an area critical for the maintenance of working memory. For this, we computed inter-area phase synchrony between the contact points in the SMC and those in the TL during switch trials after having aligned the two sets of signals to the ERP event. Interestingly, our analysis reveals synchronization between SMA and TL in the theta range, being synchrony significantly higher in fast trials compared to slow trials (fig. 4.4-D) (Z-statistics permutation ($N = 1000$) analysis: S1, $z > 1.97$, $p < 0.01$; S2, $z > 1.97$, $p < 0.01$). The subset of frequency in theta is consistent with those emerged from the inter-trial phase analysis within each patient, however, the peak of synchrony between areas follows the phase reset in the SMC. This result supports our hypothesis that deliberate action switch could involve a network of multiple regions where SMC and TM transiently couple to facilitate faster retrieval of a ”deliberate” command.

[ht]

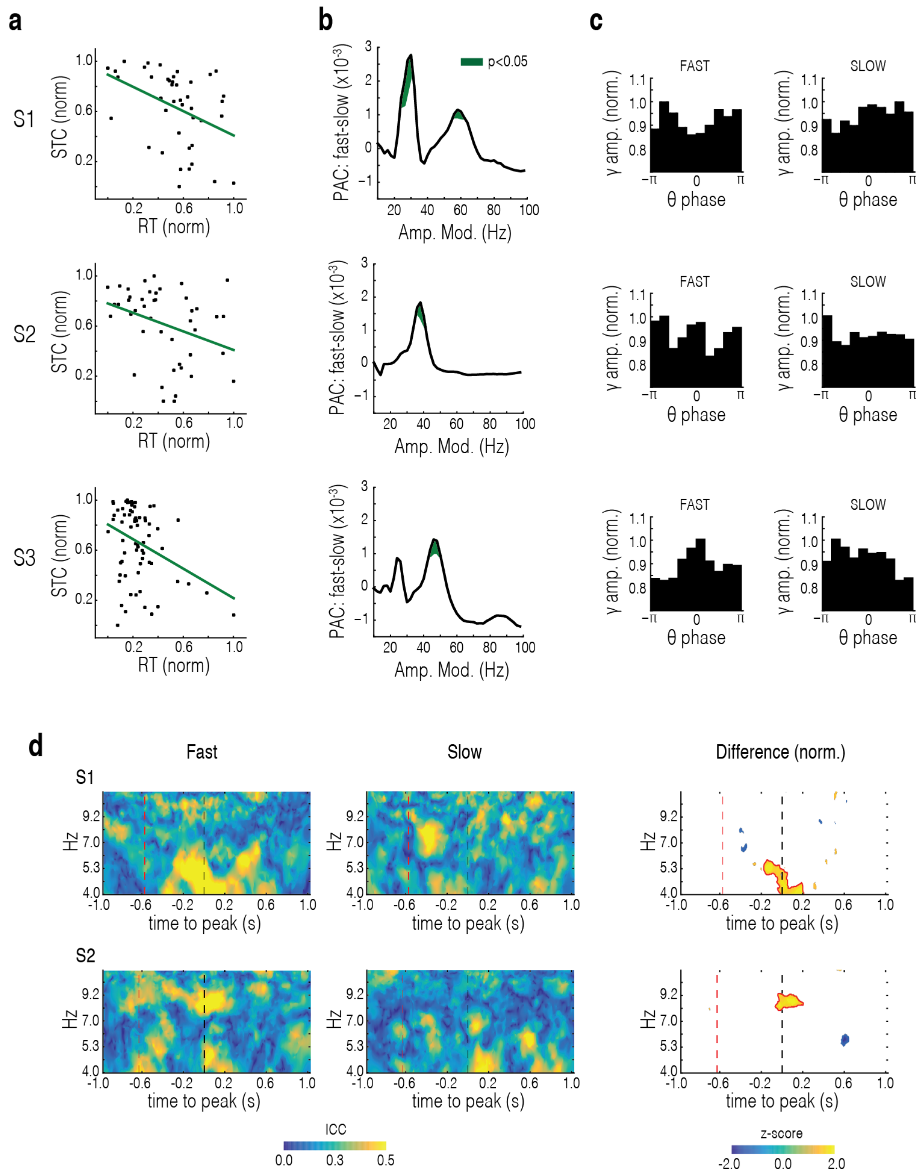


Figure 4.4: **Single trial phase coherence and phase-amplitude coupling during switch trials. a.** Relationship between Single Trial Coherence (STC) (mean of significant clusters from class comparison) and trial re-

sponse time during switch trials for each patient. Green line indicates the linear fit of the data. **b.** Phase Amplitude Coupling (PAC) between theta (5-7 Hz) phase and the amplitude of higher frequency bands expressed as the difference between fast and slow trials. Green solid area indicates $p < 0.05$. **c.** Distribution of gamma band amplitude (selected from statistically significant regions) over the phase of theta during fast (left) and slow (right) trials for each subject. **d.** Phase synchrony between SMC and TL (inter-area phase coherence) in fast (left) and slow (center) trials aligned to the peak of the ERP (black dashed) for each subject and normalized difference between trials (right). Red circle indicates $p < 0.01$.

4.3 Discussion

Switching from automatic to goal-oriented behavior according to contextual changes is a key aspect of human cognitive control. Previous studies have supported the role of the medial frontal cortex in orchestrating between these two modes [Nachev et al., 2008, Passingham et al., 2010, Hikosaka and Isoda, 2010a]. The mechanisms underlying the ability to switch between them, however, are not fully understood. Here we attempted to elucidate the mechanism through which the human SMC facilitates the execution of deliberate actions in a human intracranial setup using a variation of the Serial Reaction Time Task (SRTT).

The serial reaction time task [Nissen and Bullemer, 1987] tests the gradual shift from explicit to implicit processes in the control of action during the production of a repetitive sequence of key presses. Explicit processes drive action on the basis of sensory feedback, such as a cue indicating the key to be pressed, whereas implicit processes act in anticipation on the basis of a prediction encoding the position of the next cue [Moisello et al., 2009, Robertson, 2007, Keele et al., 2003]. The progressive shift from explicit feedback processes to implicit feed-forward control is measured by the gradual decrease of the time to press the next key in the sequence. Once the sequence is learned, the movement is faster and more stereotyped as an indicator of habituation [Sakai et al., 2004]. Consistently with previous results, in the training phase of our task, we observed a progressive and asymptotic decrease of total sequence production time, together with faster and more stereotyped IKIs in the late trials compared to initial ones, indicating that subjects correctly learned the motor sequence and relied less on explicit control. In our design, however, we extended the SRTT and introduced a switch cue that signaled a change of context and required the subjects to execute an alternative uncued key-press. We found a significant increase in execution time between a habitual and a switch key-press, which suggests that this manipulation induced a sudden change from automatic to controlled processes [Isoda and Hikosaka, 2007, Nachev et al., 2005, Crone et al., 2006].

From a theoretical perspective, SMC could be in charge of facilitating controlled actions by inhibiting ongoing behavior upon a change of context [Hikosaka and Isoda, 2010a]. In particular [Isoda and Hikosaka, 2007] showed a lateralized increase in the firing rate of primate supplementary eye field neurons that accounted for successful performance exclusively during switch but not habitual actions, suggesting an involvement of the SMC in the control of deliberate movements.

Similarly, we have shown that in humans an increase in the LFP amplitude in the medial part of Broadman area 6 could successfully predict whether the action was performed in a controlled or automatic way, supporting the role of human SMC in the deliberate action switch. Due to limitations of our experimental setup, however, we could not test the lateralization of this response, and therefore we cannot support its limb specificity. Nevertheless, the relevance of this neural event for behavior and its link to motor control was supported by the strong relationship between the single trial time to peak and the switch key-press time, whereas the lag from the peak to movement execution was relatively stereotyped suggesting little interference between this neural signature and action.

Critically, once the individual trials were aligned to the peak of the ERP, we failed to encounter a reliable predictor of the deliberate performance in the amplitude of the response. This is at odds with previous reports from human EEG showing a direct link between the amplitude of the P3 signature in the frontal areas and reaction times, where greater amplitude predicts faster responses [Ramchurn et al., 2014, Delorme et al., 2007]. This difference may be due to the intrinsic differences between EEG and intra-cranial methods, as the former suffers artifacts due to the scalp diffusion whereas the latter provides a direct recording of the neural activity from the neural tissue. The observation that amplitude of the evoked response did not relate with the response time suggested that the mechanisms underlying action switch could be encoded by other features of the neural signal such as oscillatory dynamics.

An important issue for neuroscience is to determine the functional role of oscillations in the human brain [Fries, 2015, Helfrich and Knight, 2016]. Importantly, we have shown that the phase coherence of the neural activity before deliberate action execution could predict the response time of switch trials. In particular, we found that the oscillations in the theta range were more aligned in fast than slow trials. From a methodological perspective, this result emerged by aligning the neural signals to an endogenous event (ERP peak) rather than an exogenous one and remarks the necessity to take into account the timings of internal neural dynamics, instead of artificially imposed ones, to uncover otherwise hidden phenomena [Voytek et al., 2015]. In addition, the single trial analysis revealed a negative relationship between ITPC and behavioral performance precluded to group average analysis [Cohen and Cavanagh, 2011], suggesting a direct effect of oscillatory phase on behavior.

Overall, these observations suggest that deliberate action switch in the human brain could be facilitated by a low-frequency phase code. The notion that phase-coding could represent an operational mode of the brain is supported by numerous experimental [Buzsáki and Draguhn, 2004, Engel et al., 2001] and theoretical [Fries, 2015, Lisman and Jensen, 2013] accounts. In particular, the phase of low-frequency oscillations has been linked to memory encoding processes [Siegel et al., 2009, Sweeney-Reed et al., 2014], attentional [Voloh et al., 2015, Voytek et al., 2010] and executive functions [Voytek et al., 2015, Helfrich and Knight, 2016]. Few reports have linked phase coding to behavioral performance, as in [Szczepanski et al., 2014] where the strength of PAC in frontal and parietal regions correlated with reaction times in an attentional task.

However, whether a phase coding scheme also underlies motor functions is still an open question. Here, we support the view that theta oscillations could be involved in controlling behavior under the demands of executive functions in motor areas. Similar to what observed in [Voloh et al., 2015] phase resets could be a signature of local neural populations aligning spatiotemporal dynamics in a task-relevant manner to promote prompt action execution, where greater synchrony translates to a faster encoding of a command to downstream areas (i.e. motor cortex) [Ohara et al., 2001]. An observation further supported in our data by the greater modulatory effect of theta phase on the amplitude of high-frequency oscillations during fast compared to slow trials.

In addition, phase alignment could also represent a mechanism through which the functional brain networks involved in executive functions communicate to successfully modulate behavior [Hyafil et al., 2015]. We support this hypothesis by showing that the SMC synchronizes in the theta range with the temporal lobe. Here the peak of synchronization followed the local theta phase reset in the SMC and, crucially, it was significantly related to the behavioral performance, with faster trials showing higher synchronization than slower trials. This pattern of coherent activity between these areas could reflect a mechanism through which the medial frontal cortex probes the memory system in the retrieval of the appropriate behavioral pattern to be executed in a situation when alternative behavioral objectives are not explicitly cued (i.e. uncued key).

It has been suggested that executive functions rely on frontal working-memory systems (i.e. DLPFC) for the maintenance of relevant contextual information and the rules that should drive behavior [Koechlin et al., 2003]. Although the involvement of prefrontal memory networks in deliberate action execution is likely, the temporal lobe may also give an important contribution for long-term encoding and retrieval of elements when attention is diverted or during tasks with long delays [Jeneson and Squire, 2012,Ranganath and Blumenfeld, 2005]. This might perhaps be the case in our design, where switch trials are separated by a number of seconds and the rule has to be retrieved at every trial. This observation is in agreement with previous results that suggest synchronous activity between temporal and frontal areas during a decision-making task in the presence of a memory-guided contextual rule, despite that this activity pattern does not predict performance [Guitart-Masip et al., 2013].

This result extends previous reports that suggested a role of theta frequency phase as an inter-regional communication mechanism during cognitive tasks. On the one hand, this modulation may mediate the interaction between MFC and the sensory areas such as the Parietal cortex signaling attentional demands. For example, [Phillips et al., 2014] reported greater theta synchronization between SMA and Parietal cortex in monkeys during deliberate action switch, even though no significant interaction with the behavioral performance was found. On the other hand, theta synchrony and theta-gamma CFC code have been found as a signature of the hierarchical control from pre-frontal areas over the motor system [Voytek et al., 2015]. Altogether, this and the previous results draw a scenario where executive functions are not the hallmark of one brain region, but instead, they involve a larger brain network of sensory, memory, cognitive and motor regions.

Importantly, it has been suggested that medial-frontal theta oscillations may also mediate inhibitory control by promoting action-slowing during situations of conflict and error [Cavanagh and Frank, 2014]. In particular, human EEG studies testing interference tasks reported increased power in the theta range during high-conflict trials to correlate with an increase in response time perspectively and retrospectively (i.e. post-error slowing) [Cohen and Cavanagh, 2011,Cavanagh et al., 2010,Narayanan et al., 2013]. Even though we did not explicitly test for this aspect within our paradigm, we cannot confirm the role of theta as a signature of inhibition for two reasons. Firstly, contrasting with the ’frontal theta’ interpretation, our analysis did not reveal any distinctive role of the oscillatory power of this frequency band. In addition, theta oscillatory phase alignment was found to promote fast action rather than increasing the response time. Following these observations, we propose that theta oscillations may have different functional roles captured by different features of the neural code, where power may locally encode a general conflict signal whereas phase may serve as a long-range communication channel that promotes fast executive control [Cohen, 2014b].

In conclusion, we propose an account of the role of human SMA in switching from automatic to controlled processes facilitating deliberate action execution. Automatic processes could be implemented in motor cortical and subcortical systems such as the motor cortex and the cerebellum [Sakai et al., 2002]. SMA, in turn, could be involved in monitoring action and mediating deliberate processes through a phase and CFC coding that, by integrating memory retrieval, sensory signals [Phillips et al., 2014] and goal-directed control [Voytek et al., 2010], could halt the ongoing behavior through synchronization with the sub-thalamic nucleus via the hyper-direct pathway [Zavala et al., 2016, Isoda and Tanji, 2004] and overwrite automatic actions with controlled ones in the motor cortex [Ohara et al., 2001].

4.4 Methods

4.4.1 Data collection

Data were collected from three right-handed subjects with intractable epilepsy, temporarily implanted with intracranial electrodes (iEEG) as a part of a pre-operation procedure to localize the seizure focus. Electrode placement was determined by the surgeons based on the clinical need of each patient. Data were recorded at the Epilepsy Monitoring Unit of the Hospital del Mar, Barcelona, Spain. All subjects provided the informed consent to participate in the study in accordance with the ethical committee of the Pompeu Fabra University as well as Hospital del Mar. All iEEG recordings were performed using a standard clinical EEG system (XLTEK, subsidiary of Natus Medical) with a 500 Hz sampling rate. A uni- or bilateral implantation was performed using 12 to 16 intracerebral electrodes (Dixi MÃ©dical, BesanÃ§on, France; diameter: 0.8 mm; 5 to 15 contacts, 2 mm long, 1.5 mm apart) that were stereotactically inserted using robotic guidance (ROSA, Medtech Surgical, Inc).

To identify the anatomical position of the electrode contacts we used the 3D Slicer software [Fedorov et al., 2012]. With the registration tool, we coregistered (rigid body, 6 degrees of freedom) the post-implantation CT scan to the pre-implantation MRI. We then added the electrode fiducials on a glass model of each patient's brain obtained with the segmentation tool of the Freesurfer bundle [Fischl, 2012]. To obtain a single model we coregistered all studies on the MNI152 template provided by the Freesurfer bundle using a semi-automated registration process of 3D Slicer. Briefly, we calculated a linear transform with 12 degrees of freedom by superposing and morphing each patient's brain MRI onto the MNI brain template, then we used the transform matrix to translate, shift, skew and resize all other studies (CT scan, and unaltered MRI) accordingly. Since the 3D Slicer interface shows the MNI coordinates when hovering the mouse pointer, we could identify structures touched by electrode contacts both by visual inspection and by referring to the aforementioned coordinates.

4.4.2 Behavioral task

The behavioral task was a variation of the standard Serial Reaction Time Task (SRTT), a type of paradigm that promotes automation of sequential motor behavior [Nissen and Bullemer, 1987]. Differently from the original task, however, here, in a small subset of trials, the sequential automated action was occasionally interrupted by a cue that required the subjects to switch to a different goal instructed at the beginning of the experiment. The task comprised a maximum of 500 experimental trials preceded by 60 trials of training. There were two types of trials: habitual and switch. Every trial started with a waiting period of 700 ms ± 200 ms during which the screen remained blank. After this, subjects were presented with a virtual 4 by 4 square keyboard. During habitual trials, a sequence of five keys was highlighted sequentially (green cue) upon button-press. Subjects were instructed to press the cued key as rapidly as possible until the end of the sequence. Each trial terminated at the end of the sequence, and the following one started. The sequence was pseudorandomly generated at the beginning of the experiment to respect a spatial uniform distribution over the keyboard and it was maintained constant throughout the experiment. Switch trials started with the same highlighted key as the habitual trials (green cue), and the next step in the sequence was highlighted upon a button press. Differently from habitual trials, however, one of the intermediate steps of the sequence (i.e. step 2-4 selected at random) highlighted in red (switch cue). Upon presentation of the switch cue, subjects were required to halt the ongoing sequence of movements as fast as possible and press an alternative, uncued key. Participants received all the instructions prior to the beginning of the experiment. Feedback was provided for neither the correct nor incorrect performance. The training phase only comprised habitual trials, whereas the experimental phase included a combination of habitual and switch trials pseudo-randomly interspersed every 7 ± 2 trials. The experimental setup ran on a portable capacitive screen fixed to the hospital overbed-table. The tablet included a custom-made Java-based application running the experimental task and logged behavioral performance at 50 Hz whereas task synchronization with the neural recordings was achieved through serial communication with the recording system. Subjects sat in a comfortable position that avoided motor constraints to the arm. After receiving the instructions, subjects underwent a short session that exemplified the task. After this, the experimental session started. Subjects could withdraw at any point during the task.

4.4.3 Electrophysiology pre-processing

All electrophysiological data were preprocessed in Matlab (EEGLAB toolbox) and subsequently analyzed in Python using custom scripts based on the Numpy , Scipy , SkLearn and MNE libraries. Data were initially filtered using a two-way zero phase-lag, FIR bandpass filter (2-200 HZ) and an additional notch filter (window = 2Hz) at 50Hz, 100Hz and 150Hz to remove AC current contamination and respective harmonics. Following this step, the signals were individually re-referenced to the average potential of all electrodes for each subject. After filtering, artifacts derived from strong muscle activity or interference due to contact with electrical devices were identified by visual inspection and respective epochs rejected. To reduce remaining artifacts (i.e. cardiac artifacts, muscle twitches), we applied a combination of Principal Component Analysis (PCA) and independent component analysis (ICA). In brief, we performed PCA on all channels and identified those components which accounted for $> 98\%$ of the variance. Such components were subsequently decomposed into the same number of independent components through ICA. At this point, each component time-series was visually inspected and components that reflected signal artifacts were rejected. The selection of artifact components was based on a careful inspection of their power spectrum, correlation with other physiological measures (i.e. ECG), and the relation to the temporal structure of the experiment.

4.4.4 Amplitude analysis

For each subject, the filtered and artifact-free signal was split into epochs according to the trial structure of the task. Each epoch was individually base-line corrected by subtracting the mean amplitude value in a temporal window of 500 ms preceding the beginning of each trial. To identify task-selective channels displaying changes in the amplitude of the signal (i.e. Event-Related Potentials (ERP)) we extracted a set of 3 descriptors (absolute mean, variance and integral) and applied a classification method based on the Linear Discriminant Analysis (LDA) [Blankertz et al., 2011]. 100 cross-validation steps were performed to assess performance with Fishers F1 score on class-balanced bootstraps of data samples (80% training, 20% testing). The channels providing the highest classification accuracy were finally selected as the task-related channels. Note that this analysis was naive with regards to the electrode location or the polarity of the event. This step allowed us to narrow down our analysis to those contact points that displayed a task-related change in the amplitude (a detectable difference between conditions) for each subject.

Spectral analysis revealed the presence of ERPs in the low-frequency range between 2 and 4 HZ. Trial-by-trial ERP peaks in the switch condition were therefore identified by band-passing the signal in the 2-4 Hz range using a two-ways zero-phase FIR filter and applying a peak detection algorithm that estimated the time of the absolute peak amplitude between stimulus presentation (switch-cue) and the response. Single-trial stimulus-peak interval, as well as peak-response interval, were further calculated by subtracting the stimulus presentation time from the peak time and the peak time from the response time respectively.

Finally, the statistical analysis of amplitude differences was performed through a T-statistics one-dimensional non-parametric cluster based permutation test [Maris and Oostenveld, 2007] as implemented in the MNE toolbox with cluster significance threshold = 0.05 and number of permutations = 1000.

4.4.5 Spectral Analysis

Spectral analyses were performed using a DPSS multitaper method [Mitra and Pesaran, 1999, Thomson, 1982] as implemented in the MNI toolbox.

Changes in the power with respect to the baseline where computed by z-transforming the power spectrum . Statistical differences in the time-frequency power between conditions were calculated through T-statistics two-dimensional non-parametric cluster based permutation analysis as implemented in the MNE toolbox setting cluster significance threshold = 0.05 and number of permutations = 1000 [Maris and Oostenveld, 2007].

Inter-trial phase coherence (ITPC)

We estimated inter-trial phase coherence to quantify the frequency-dependent synchronization across trials through Phase Locking Value (PLV) method [Lachaux et al., 1999]. ITCP is computed as:

$$ITPC = \frac{1}{N} | \sum_{n=1}^{N} e^{j\phi_n} |$$

where N is the number of trials in one condition and ϕ_n represents the phase estimate at the n_{th} trial. ITCP is bounded between 0 and 1, where 1 represents full phase synchronization. In order to test differences in ITCP between conditions, we used the cluster-based permutations method proposed by [Maris et al., 2007]. First, we applied a z-transform to the difference in coherence between conditions (Z_{ITPC}) that rendered the distribution approximately normal [Enochson and Goodman, 1965]:

$$Z_{ITPC} = \frac{(tanh^{-1}(|ITPC_1(f)|) - (1/d.f._1 - 2)) - (tanh^{-1}(|ITPC_2(f)|) - (1/d.f._2 - 2))}{\sqrt{(1/d.f._1 - 2) + (1/d.f._2 - 2)}}$$

To account for the positive bias of ITPC, we used the same amount of trials for the two conditions compared. Second, we selected those regions where $z > 2.58$ corresponding to the $99th$ percentile of the distribution. Finally, we assessed the significance of the measured difference against the H0 obtained by computing the coherence difference between surrogate groups constructed by permuting 1000 times the original labels and extracting the resulting Montecarlo P value.

Single trial ITPC

ITPC is by definition an average measure across multiple trials. An estimate of the contribution of the single trial to the average ITPC (STPC), however, can be obtained by computing the difference between the ITPC across all trials and the ITPC across all but one trial following the method proposed by [Jarvis and Mitra, 2001] and previously applied by [Hipp et al., 2011]. The Single Trial ITPC ($STPC_i$) for the ith trial is computed as follows:

$$STPC_i = N \cdot ITPC_z^{all} - (N-1) \cdot ITPC_z^{all-i}$$

where N is the number of trials and $ITPC_z^{all}$ and $ITPC_z^{all-i}$ are the z-transformed ITPCs for all trials and all but the ith trial respectively. Finally, to obtain a single trial value of STPC we selected as Region Of Interest (ROI) the time-frequency points of the spectrum that resulted significant from the cluster based permutation analysis and integrated the STPC values for that ROI.

4.4.6 Phase-amplitude coupling (PAC)

PAC is a measure that quantifies the modulatory effect of low-frequency phase on higher frequency amplitude as a signature of the interaction between their underlying processes resonating at different frequency bands. PAC was computed through the Generalized Linear Models (GLM) method [Penny et al., 2008] that captures the proportion of variance explained by an underlying linear relationship between analytical amplitude (i.e. envelope, modulated) and phase (modulating) as obtained by Hilbert transforming the signal, using the PACpy toolbox. We restricted our analysis of PAC to the ROIs emerged from cluster based permutation analysis and selected as modulatory frequency band the significant frequency domain range for each patient. Our epoch selection was also restricted to the temporal window where a significant increase in phase alignment was detected. For each subject, we obtained one surrogate signal for fast and slow trials by concatenating the respective single trial windows, so to achieve the temporal resolution necessary for this type of analysis. Further, we computed PAC values between the selected modulatory phase and the amplitude of higher frequencies (10-100 Hz in steps of 2 Hz, modulated frequency), and obtained the difference between the two conditions. Statistical significance between the two conditions was tested through z-statistics against the null-hypothesis of samples from both conditions belonging to the same distribution. This was obtained by randomly permuting the conditions’ labels and calculating the 95 percentile of the maximum PAC value achieved under the assumption that the two conditions were sampled from the same distribution.

Chapter 5

CONCLUSIONS

The aim of this dissertation was to advance our understanding on the multifaceted way the brain controls behavior. We focused our approach on the definitions of automatic and deliberate processes and mapped them to their computational and neurophysiological underpinnings with the goal to explain their role in regulating the interaction of an agent within its environment. In doing this, we have contributed to elucidate some of the main system-level interactions underlying these processes from a theoretical perspective and demonstrated their relevance in controlling behavior. Further, we generated specific individual hypotheses on some of the key neural substrates that underlie automatic and deliberate control. On the one hand, starting from the formulation of an anticipatory control problem we reformulated the traditional motor-centric role of the cerebellum within the sensory domain and advanced new theoretical insights on its functions, together with concrete experimental predictions. On the other hand, we investigated the active role of the human medial frontal cortex in controlling the switch between automatic and deliberate processes, supporting the system level interactions between goals and the memory system hypothesized in the chapter 2. Throughout the experimental chapters of this dissertation, we combined a number of methods, from computational, to robotic and electrophysiological, where each one of these approaches contributed with a distinctive insight. Computational modeling provided theoretical predictions informed by biology on the mechanisms and system-level interactions underlying automatic and deliberate behavior. Robotics imposed the real-world constrains necessary for the analysis of the interaction between an embodied system and its environment. Finally, neurophysiology constituted the ground for the empirical validation of theoretical insights on brain functions. This variegated methodological approach implicitly emphasized the need to look at the brain from multiple levels of description to capture the complexity of biological behavior.

The coexistence of automatic and deliberate processes is central to a number of theories that stem from the analysis of behavior of animals and humans. Their characterization however has mainly focused on their information processing. This perspective is perhaps grounded in the cognitivist tradition whose primary goal is to study the way the brain and the mind acquire, process, and retrieve information in order to build an accurate representation of the world. This approach is fundamental in order to advance theoretical interpretations on the local computations of individual brain circuits, and it can provide useful insights on the learning mechanisms that underlie animal performance. Nevertheless, we have argued that an approach focusing only on the brain processing functions may overlook an important aspect: the brain evolved to control a physical body and regulate its interactions with the environment. Integrating, therefore, a view on the brain that takes into account the control problems which an embodied system needs to solve in order to perceive and act within the environment with the goal of survival can lead to the understanding of the functional role of automatic and deliberate processes as well as their organization. For this reason, in chapter 2 of this dissertation, we aimed at bringing an integrated view on some of the key brain areas that underlie automatic and controlled processes from an embodied point of view. We did this through the formulation of a biologically constrained control system based on the anatomy and physiology of the rodent brain that advanced a concrete definition of the computational requirements that the brain as a control system has to implement in order to produce complex behavior. In doing so, we provided a bridge between the information processing perspective, intrinsic to the implementation and analysis of individual modules, and the control perspective emerging from the global objective of controlling the behavior of an artifact in the physical world. In particular, three insights have emerged from our first contribution. Firstly, we have shown that while individual modules process information through distinct computational mechanisms, their integration serves the control of behavior. This emerged in the simple sensory-motor mappings provided by the reactive system that contributed to the early stages of learning laying the ground for exploratory behaviors. It also emerged within the associative learning mechanisms that learned to associate stimuli and responses to predictively react to behaviorally relevant cues, and deliberate mechanisms that acquired an internal representation that sub-served goal-oriented behavior and planning. Importantly, each layer contained distinct modules implementing different learning mechanisms but no global objective function was defined within the system. Nevertheless, the agent tended towards a general optimization of behavior. This suggests that increasingly sophisticated behavioral strategies can be achieved implicitly as an emergent property of the interaction between multiple learning mechanisms rather than by explicitly optimizing one objective function. This is a relevant observation with respect to the normative frameworks such as reinforcement learning or Bayesian inference since it offers a way to reconcile the diversity of brain structures and computation with the global tendency towards optimal behavior. A second point which emerged from the first contribution is that, by looking at the progressive development of different integrated learning mechanisms from reactive to adaptive and contextual, we have provided a perspective on automatic and deliberate control not as two separated processes but rather as a continuum, where each form of control builds hierarchically on top of the lower one. Reactive mechanisms build on somatic features of the agent such as sensors and actuators. Associative learning is based on reactive mechanisms, where reflexes are progressively advanced through the pairing with predictive cues. Finally, deliberate processes extend the adaptive properties of the automatic system from the short-time scale (i.e. S-R) to the long one by using an internal representation that stems from the situated interactions provided by lower layers. Indeed, we have shown how the progressive enabling of these three layers of control lead to increased optimality. This could rise the question of why the brain does not just have a deliberate system? The answer might be that it would not be possible since enabling the deliberate system alone would lack the grounding for learning and representation. Indeed, a careful analysis of the quality of the information processed by every module of the system unveils its sensory-motor and embodied nature. The reactive system converts sensory inputs into reflexive responses that allow for random navigation. The cerebellum extends the actions of the reactive system to novel contingencies through sensory-motor associative learning. The hippocampus builds internal representations of the environment through the integration of speed, head direction and contingent external cues, that is, features extracted from the embodied interaction with the environment. Finally, the decision-making system extracts and orchestrates goals that are representation of low-level needs associated to the internal bodily state. This is consistent with an evolutionary and developmental perspective on brain anatomy and functions that identifies the structures responsible for reflexive behavior as philogenetically oldest and anatomically closer to the spinal cord [Jerison, 2012]. In turn, associative mechanisms have emerged later in the evolution of the brain, but in tight anatomical relationship with reactive ones. Areas related to deliberate control that implements internal representations such as the hippocampus and the neocortex are philogenetically newest and found principally in mammals. This perspective is possible only by looking at the evolution of learning in an embodied agent within its environment and it would not be possible by taking a sole (disembodied) information processing perspective, as this would lack the sensory-motor foundations of learning. Related to the previous points is the insight that automatic and deliberate processes are highly distributed and cannot be reconciled with the functions of one brain area or one individual processor and that a mapping between functional level of description and neural substrate is necessary to understand the neural mechanism underling behavior. For this, a computational implementation of this mapping into an embodied controlled system offers a valuable analytical tool to acquire insights and make testable hypotheses on the design principles that various controllers in the brain may implement. We have expanded some of these insights in the following chapters of the thesis.

We have grounded our definition of an automatic system in terms of classical conditioning of avoidant responses in the cerebellum. In chapter 1, we defined the automatic control system as the interaction between reactive and adaptive layers of control. Here, multiple parallel microcircuits locally implemented mechanisms that reflect the internal information processing of the cerebellum and that underlie the associative learning found in avoidance learning. From a control perspective this structure learned to produce anticipatory responses to environmental stimuli, which improved the motor control capabilities of the agent. In particular, in contextualizing conditioning phenomena within an embodied system, anticipatory responses provided a form of feedforward control that complemented reactive, reflex-like behaviors. The use of classical conditioning paradigms in robots embedding biologically plausible models of the cerebellum within their control architectures has a long-standing tradition [Van Der Smagt, 1998, HofstoÈtter et al., 2002, Herreros and Verschure, 2013a, McKinstry et al., 2006]. However, while these studies have typically focused on the analysis of one microcircuit, our contribution was to integrate a number of microcircuits and show how the parallel feedforward control scheme can enhance the sensory-motor capabilities of an agent. From a cerebellar learning perspective, this is particularly relevant. The parallel anatomical organization of the cerebellum indeed suggests an homogeneous processing structure whose behavioral functions are determined by its input-output connectivity. However, this assumption has been recently challenged by recent discoveries suggesting a richer arrangement of anatomical connections that may justify alternative, perhaps sequential, configurations [Apps and Garwicz, 2005, Cerminara et al., 2015].

We have explored the possibility of alternative schemes of cerebellar learning and their role in controlling anticipatory behavior in part 2. In doing this, we have focused on the acquisition of anticipatory postural adjustments, a form of predictive control dependent on the cerebellum that allows the minimization of externally induced displacements through the anticipation of a motor command. Following the modeling work presented in part 1, we hypothesized that postural adjustments could be the result of parallel sensory-motor associations acquired by a control scheme based on inverse models. Nevertheless, we observed that from a control perspective this formulation lacked the robustness to deal with uncertainty. We hypothesized that in order to overcome this limitation the cerebellum could learn a sensory prediction, namely an associative representation between two stimuli (chapter 3). This hypothesis on the cerebellar functions let to the reformulation of the adaptive control architecture presented in chapter 3 (HSPC).

In the HSPC architecture sensory predictions are organized in a hierarchical structure that mirrors the cause-effect relationship within the task. This is achieved by serially linking multiple microcircuits that acquire qualitatively different sensory predictions. The anticipatory module of HSPC, in particular, learns the association between one sensory input (cue) and another one (impact). This form of learning is consistent with the Rescorla-Wagner class of interpretations of conditioning, where the learned relationship between two stimuli allows to make predictions about future states (see [Balkenius and Morén, 1998] for review). In HSPC, however, the prediction of a sensory event is not just an acquired knowledge of the environment but it serves the purpose of control. This aspect emerges from the temporal dynamics of the prediction, and in particular, it is encoded in the delay that determines the amount of anticipation that will be propagated to downstream modules in order to drive an effective predictive response. The degree of anticipation needs to be tuned to the dynamics of the body in its interaction with the physical world. Importantly, temporal dynamics of predicted errors are also central to the feed-forward compensatory module of HSPC and that extends the model of cerebellar learning previously advances by the CFPC architecture [Herreros and Verschure, 2015]. Indeed in CFPC the system learns to predict a counterfactual error with a temporal accuracy matched to behavioral functions [Suvrathan et al., 2016]. Altogether, predictions in HSPC encode an action-aware knowledge about the dynamical relationship between perceptual events that has the goal to minimize errors in the sensory domain through action. As discussed more in detail in chapter 3, HSPC follows a scheme consistent with the theory of active inference. In active inference, indeed, a hierarchical organization of descending sensory predictions encoding an internal (generative) representation of perceptual modalities have the goal to minimize ascending prediction errors by actively modifying perception [Friston, 2011a]. In sum, the work presented in part 2 contributes to the understanding of motor learning proposing a novel control scheme that can account for acquisition, generalization and robustness in anticipatory control paradigms. It provides an interpretation of associative learning that reconciles the causal knowledge view of conditioning with the notion of predictive control and shows how internal representations can be used for behavioral purposes. In addition, HSPC bridges between the normative framework of active inference and a biologically plausible scheme based on the computation of the cerebellum. This proposal has concrete implications for the theory of the cerebellum. Firstly, it could contribute to reconcile two distinct views on the cerebellum in motor and non-motor domains. Indeed, HSPC is consistent with the motor control literature as it makes use of internal models that act as feed-forward controllers to drive motor behavior [Wolpert et al., 1998b]. However, unlike the traditional inverse models it does this in the sensory domain, and unlike forward models it can anticipate sensory consequences that are not contingent upon motor commands. In this, HSPC goes beyond the traditional distinction between forward and inverse models and advances a hypothesis on the existence of ”forward models of the environment”. Forward models of the environment, namely internal models of sensory events timed to behavioral goals may explain the involvement of the cerebellum in the prediction of temporal aspects of perceptual stimuli [Roth et al., 2013a, Deluca et al., 2014, Therrien and Bastian, 2015], further used to drive behavior (i.e. motor preparation) [Ramnani and Passingham, 2001, Sakai et al., 2002]. A second implication following our proposal is that cerebellar predictions could be organized hierarchically. We have speculated that a sequential organization of cerebellar circuits could be implemented by the directional organization of the nucleo-cortical projections between the Nucleus Interpositus Posterior (NIP) and the Nucleus Interpositus Anterior (NIA) [Apps and Garwicz, 2005]. However, a sequential organization of cerebellar predictions could also be justified by the arrangements of cortico-cerebellar projections [Kelly and Strick, 2003]. In particular, the cerebellum is anatomically and functionally connected with the parietal cortex, as well as motor, premotor and frontal areas involved in processing perception and action at multiple levels of a functional hierarchy. It is therefore possible that the cerebellum could participate at each level of this hierarchy, or gradient [Guell et al., 2018], by advancing qualitatively distinct predictions used to modulate behavior [Ramnani and Passingham, 2001]. Future research in this direction should focus on testing the experimental predictions advanced by the HSPC theoretical model both through behavioral and electrophysiological methods. Behavioral experiments should focus on testing the relationship between distal and proximal events in generating anticipatory behavior. In particular, HSPC predicts a causal link between distal and proximal and vestibular modalities in the generation of anticipatory postural adjustments. Even though some literature suggests this could be the case [Mohapatra et al., 2012,Stapley et al., 2002], further experiments concretely addressing this question should be designed, for example, by decoupling distal and proximal cues (i.e. catch trial) in a virtual reality postural task. However, we expect the underlying scheme presented for postural control to generalize to other anticipatory responses (i.e. limb). This could be tested in an fMRI setup to elucidate the contribution of the areas involved in the acquisition of anticipatory responses, and crucially, those involved in the computation of errors. Here, HSPC would predict a high involvement of the cerebellum during acquisition of sensory predictions and the somatosensory areas in detecting sensory-prediction errors [Mathis et al., 2017]. From an electrophysiological perspective, future research should focus on the identification of functional hierarchies between distinct zones of the cerebellum, for example by simultaneously recording from NIP and NIA in order to quantify the amount of directional information transferred from one zone to the next (i.e. granger causality). Further work on the control aspects of HSPC should be also be addressed. Specifically, HSPC implicitly suggests that sensory predictions broaden the limits of control allowed to the feedback controller (i.e. the range of disturbances in which it can operate) while maintaining robustness. A systematic quantification of this trade-off in comparison with standard solutions could lead to the design of novel control architectures that make strong use of feedforward control such as those used in soft robotics.

We have grounded our definition of deliberate control within the frontal cortex. In chapter 2 of this dissertation, we have suggested that one of the key differences between automatic and deliberate processes is the fact that the automatic system, in its associative nature, does not rely on explicit internal representations of behavioral goals. Within the automatic system the behavioral goal was set by the intrinsic role of the reflex, representing a pre-wired behavioral repertoire (i.e. avoiding an obstacle). However, in order to achieve flexible goal oriented behavior a system requires an internal representation that allows action planning. We have shown in Part 1 that the deliberate system could implement the distributed functions of planning through the interaction of frontal (decision-making) areas, the memory system within the hippocampus and the basal ganglia. In particular, goals were encoded in frontal circuits that converted low level drives into higher-level decision-making (i.e. what goal to pursue). A memory of the environment was stored in hippocampal place cells that progressively processed streams of sensory inputs into unique internal representations. A long-term memory structure, in turn, stored representations of locations and resources once those were encountered. Finally, the motor system was in charge of implementing goal-oriented actions through the action selection mechanisms found in the basal ganglia. This is consistent with the notion of model based planning where an acquired schema of state (location in space), action (behavioral repertoire) and value (type of resources present in the environment) is used to select an appropriate behavioral strategy that leads to increased efficiency of the system. Here our contribution was to bring together a biologically plausible implementation of how deliberate control of action could be implemented in the brain. Empirical evidence supports indeed the role of each individual sub-system presented within our contribution in deliberate control. For example, neural signals representing goals have been decoded in the rodent hippocampus as predictive of task choice. These signals were mediated by a fronto-hippocampal pathway possibly linking internal representations of goals and space [Ito et al., 2015]. Prefrontal circuits indeed orchestrate the selection of goals on the basis of motivational signals encoded by dopaminergic neurons in the VTA [D’Ardenne et al., 2012,Grace et al., 2007]. Finally, the basal ganglia could be involved in the control of action selection mechanism driven by deliberate circuits through inhibition of ongoing plans and facilitation of novel ones. Nevertheless, the interactions of these sub-systems and how they orchestrate goal-oriented action is still under debate. Therefore we feel that contributing to a working model on how multiple areas of the brain could combine their internal computation to produce behavior could advance theoretical and empirical studies. Indeed, an important theoretical question that followed our proposal is how the brain can convert representations of goals and strategies encoded in memory, into actions implemented in the motor system. This is of great relevance for the current discussion because making this link could allow to capture two complementary aspects of brain functions, such as the information processing capabilities intrinsic to the formation and retrieval of internal representations and the control aspects required to act in the physical world. In chapter 4, we have attempted to address this question in the context of a human neurophysiological setup. Based on previous reports, we have hypothesized that the Supplementary Motor Complex (SMC) in the medial frontal cortex could constitute a likely candidate for linking higher level planning (i.e. goals) with motor execution [Nachev et al., 2008]. In particular, following a theoretical definition informed by empirical evidence on the functions of this area in primates [Hikosaka and Isoda, 2010a], we outlined four functional requirements that the SMC should meet in order to qualify as a controller orchestrating goal-oriented actions. These functions included the ability to process relevant information and to access internal representations as well as a direct role in inhibiting current ongoing actions and facilitating alternative ones. We tested some of these aspects within a change of plans task where human subjects implanted with intracranial electrodes had to perform a stereotyped sequence of movements occasionally interrupted by a switch cue. Upon presentation of the cue, subjects had to interrupt the ongoing sequence and execute an alternative (deliberate) action. Our main finding was that, within the SMC, oscillatory phase in the theta range could predict deliberate behavioral responses. This result opens the question of what is the role of the phase of low frequency oscillations during deliberate actions and, in particular, what aspect of the action does it encode? We have proposed that phase synchrony in low oscillations could be a signature of motor preparation, where the degree of motor preparatory activity in response to the cue could directly drive behavior. However, motor preparation in this case would not be a localized phenomenon within the SMC, but rather the result of an interplay of distinct areas within a functional network. Support for this interpretation can be found in our results by the higher Phase Amplitude Coupling (PAC) between theta and gamma bands during faster trials. PAC is often detected as a signature of long-range communication between distinct areas within a functional network [Hyafil et al., 2015, Helfrich and Knight, 2016]. Therefore, according to this interpretation theta could represent a communication channel that mediates the interplay between various sub-system involved in the deliberate control of action [Fries, 2015]. Long-range communication in the theta range is known to mediate the communication within the fronto-parietal network during change of plan tasks and could represent a mean to integrate perceptual evidence to prepare for contextual changes [Phillips et al., 2014]. Theta is also found in the communication between prefrontal and premotor areas, where it is thought to be a signature of executive control over the motor system, especially when behavioral choices are based on high-order rules [Voytek et al., 2015]. Finally, we have found synchrony in the theta range between the temporal and the medial frontal cortex. This might mean that those two areas functionally synchronize under task demands. We speculate that this interaction could underlie the need for the deliberate control network to access memory representations of stimulus-action rules. An interpretation supported by the involvement of the temporal lobe in episodic memory formation and its operational mode in the theta range. What needs to be determined, however, is the pathway through which this communication could occur. It is possible that this synchronization could be mediated by the thalamus or by the fronto-temporal network. Nevertheless, further work should address in a more systematic way the involvement of the memory network in the deliberate control of action.

Indeed, our current experiment does not allow us to distinguish the content of this representation, which makes this interaction difficult to interpret. In order to disambiguate this, a more complex version of the proposed task should involve different types of switch cues leading to different types of responses driven by either perceptual cues or memory.

In sum, we have sketched a possible distributed network of several different cortical systems that may be involved in the control of deliberate action. Within this network goals, memory, and perception may be integrated through low frequency communication channels. An open question, however, remains as to how the deliberate network could directly control movements? We have suggested that the SMC should have both facilitatory and inhibitory functions over the motor system [Hikosaka and Isoda, 2010b]. Previous evidence supports the involvement of the subthalamic nucleus in the inhibitory aspects of the switch [Isoda and Tanji, 2004]. This is further confirmed by the anatomical connections from SMC to the STN as well as by a pattern of synchronization between these two areas in the theta range during conflict where higher synchronization leads to slower reaction times. The facilitatory pathway instead has not been fully characterized. The SMA could promote action within the motor system directly through direct projections to the spinal chord [Tanji, 1994] by exerting control over the primary motor cortex [Ohara et al., 2001]. Alternatively, the SMA could promote action through the direct pathway of the basal ganglia [Hikosaka and Isoda, 2010b]. Further studies will be needed to determine the output pathways of the deliberate control system, for example, by simultaneously recording from the medial frontal cortex and the caudate nucleus of the striatum. If this pathway promotes switching actions, behavioral dependent coherence in low frequency oscillations might be found. In general, we speculate that within this picture theta oscillations could have the double role of communication and control by mediating a distributed action aware representation promoting goal-oriented action. However, although converging evidence seem to point at low frequency oscillations as the communication channel that integrates the information processing and motor related aspects of the deliberate control network, it still remains to be determined what are the mechanisms that generate this stereotypical oscillatory pattern and how this relates to the neural activity. It has been proposed that theta oscillations could emerge from a conflict detection mechanism present in the layer 5 of the medial frontal cortex via the integration of input from other regions such as the amygdala and the fronto-parietal circuit coming from more superficial cortical layers. Integration of such inputs would increase the power of theta oscillations [Cohen, 2014b, Ulrich, 2002]. Alternatively, theta phase dynamics could be the result of synchronous activity at the output cortical layer (i.e. layer 6) and serve as communication pathway towards other brain circuits to modify cognitive and motor processes. To test this hypothesis, single-cell intracranial electrophysiology could shed light on the contributions of different layers of medial frontal cortex in integration, conflict detection and control.

To conclude, we provide an outlook and future work that concerns the system level interactions between the automatic and deliberate processes. Indeed, although in some cases we have stressed their complementarity, throughout this dissertation we have mostly analyzed them in isolation. A great body of evidence however has stressed the existence of anatomical pathway through which these systems can communicate and operate in synchrony. We feel that perhaps on of the most important anatomical pathways whose functions are still largely unexplored is the one between the lobule VII of the cerebellum and the frontal cortex, and in particular the area 46 [Ramnani, 2006a, Kelly and Strick, 2003]. The fronto-cerebellar system is indeed one of the least undetstood in the brain and it could represent a mean through which the deliberate and automatic systems interact. This interaction could have various functions that range from the automation of cognitive processes, such as rules [Balsters et al., 2012, Ramnani, 2014], but it could also support predictive processes contributing to inference and decision making under uncertainty [Blackwood et al., 2004]. This could be achieved by an updated model of our HSPC architecture, configured different processing stages of the deliberate system. Another important anatomical pathway that could underlie the interaction between the deliberate and automatic system is mediated by the basal ganglia. The basal ganglia indeed is involved in the action selection of deliberate motor programs and it receives major cortical inputs from frontal and motor areas. Recent anatomical evidence has stressed the existence of reciprocal pathways between the cerebellum and the basal ganglia [Hoshi et al., 2005, Bostan et al., 2010]. Interestingly, excitatory projections from the basal ganglia to the pontine nuclei come from the sub-thalamic nucleus, which has the role to inhibit ongoing motor program. Why would an inhibitory motor structure elicit alternative responses within the automatic system? Perhaps this could be a pathway for the deliberate system, and in particular the medial frontal cortex to indirectly recruit the cerebellum through the hyper direct pathway within the basal ganglia. To date however, only a few hypotheses have been advanced on the role of this interaction about their possible involvement in model-based action planning [Caligiore et al., 2017].

We propose that an interplay of interdisciplinary methods including computational modeling, robotics as well as behavioral and neurophysiological experiments is the key to advance the theory, unveil the unsolved mysteries about the brain, and to design the next generation of intelligent machines.

Bibliography

[Abarca et al., 1985] Abarca, N., Fantino, E., and Ito, M. (1985). Percentage reward in an operant analogue to foraging. *Animal Behaviour*, 33(4):1096–1101.

[Adams et al., 2013] Adams, R. A., Shipp, S., and Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Structure and Function*, 218(3):611–643.

[Agetsuma, 1998] Agetsuma, N. (1998). Simulation of patch use by monkeys using operant conditioning. *Journal of Ethology*, 16(2):49–55.

[Ahmed and Wolpert, 2009] Ahmed, A. A. and Wolpert, D. M. (2009). Transfer of dynamic learning across postures. *Journal of neurophysiology*, 102(5):2816–2824.

[Albert and Shadmehr, 2016] Albert, S. T. and Shadmehr, R. (2016). The neural feedback response to error as a teaching signal for the motor learning system. *Journal of Neuroscience*, 36(17):4832–4845.

[Albus, 1971] Albus, J. S. (1971). A theory of cerebellar function. *Mathematical Biosciences*, 10(1-2):25–61.

[Allum et al., 1998] Allum, J., Bloem, B., Carpenter, M., Hulliger, M., and Hadders-Algra, M. (1998). Proprioceptive control of posture: a review of new concepts. *Gait & posture*, 8(3):214–242.

[Anderson, 1983] Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22(3):261–295.

[Anderson et al., 2004] Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004). An Integrated Theory of the Mind. *Psychological Review*.

[Angaut, 1970] Angaut, P. (1970). The ascending projections of the nucleus interpositus posterior of the cat cerebellum: An experimental anatomical study using silver impregnation methods. *Brain Research*, 24(3):377–394.

[Apps and Garwicz, 2005] Apps, R. and Garwicz, M. (2005). Anatomical and physiological foundations of cerebellar information processing. *Nature Reviews Neuroscience*, 6(4):297.

[Aron and Poldrack, 2006] Aron, A. R. and Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. *Journal of Neuroscience*, 26(9):2424–2433.

[Arsiwalla et al., 2015] Arsiwalla, X. D., Zucca, R., Betella, A., Martinez, E., Dalmazzo, D., Omedas, P., Deco, G., and Verschure, P. F. M. J. (2015). Network dynamics with BrainX3: a large-scale simulation of the human brain network with real-time interaction. *Frontiers in Neuroinformatics*, 9.

[Ashby, 1949] Ashby, W. R. (1949). Critical review. the facts and methods of cybernetics-cybernetics, or control and communication in the animal and the machine. by prof. norbert wiener. new york: John wiley & sons. london: Chapman & hall. 1948. pp. 194. *Journal of Mental Science*, 95(400):716–724.

[Balkenius and Morén, 1998] Balkenius, C. and Morén, J. (1998). Computational Models of Classical Conditioning: A Comparative Study.

[Balsters et al., 2012] Balsters, J. H., Whelan, C. D., Robertson, I. H., and Ramnani, N. (2012). Cerebellum and cognition: evidence for the encoding of higher order rules. *Cerebral Cortex*, 23(6):1433–1443.

[Barlow, 1989] Barlow, H. B. (1989). Unsupervised learning. *Neural computation*, 1(3):295–311.

[Barnes et al., 2000] Barnes, G. R., Barnes, D., and Chakraborti, S. (2000). Ocular pursuit responses to repeated, single-cycle sinusoids reveal behavior compatible with predictive pursuit. *Journal of Neurophysiology*, 84(5):2340–2355.

[Baumann and Mattingley, 2010] Baumann, O. and Mattingley, J. B. (2010). Scaling of neural responses to visual and auditory motion in the human cerebellum. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(12):4489–95.

[Bernardet et al., 2008] Bernardet, U., Bermúdez I Badia, S., and Verschure, P. F. M. J. (2008). A model for the neuronal substrate of dead reckoning and memory in arthropods: a comparative computational and behavioral study. *Theory in biosciences = Theorie in den Biowissenschaften*, 127(2):163–75.

[Bizzi et al., 2008] Bizzi, E., Cheung, V. C. K., D'Avella, A., Saltiel, P., and Tresch, M. (2008). Combining modules for movement. *Brain research reviews*, 57(1):125–33.

[Blackwood et al., 2004] Blackwood, N., Simmons, A., Bentall, R., Murray, R., Howard, R., et al. (2004). The cerebellum and decision making under uncertainty. *Cognitive Brain Research*, 20(1):46–53.

[Blakemore et al., 1998] Blakemore, S. J., Wolpert, D. M., and Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature neuroscience*, 1(7):635–40.

[Blankertz et al., 2011] Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of erp components-a tutorial. *NeuroImage*, 56(2):814–825.

[Blouet and Schwartz, 2010] Blouet, C. and Schwartz, G. J. (2010). Hypothalamic nutrient sensing in the control of energy homeostasis. *Behavioural brain research*, 209(1):1–12.

[Blumenfeld and Ranganath, 2007] Blumenfeld, R. S. and Ranganath, C. (2007). Prefrontal cortex and long-term memory encoding: an integrative review of findings from neuropsychology and neuroimaging. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, 13(3):280–91.

[Bostan et al., 2010] Bostan, A. C., Dum, R. P., and Strick, P. L. (2010). The basal ganglia communicate with the cerebellum. *Proceedings of the national academy of sciences*, 107(18):8452–8456.

[Bracha et al., 1991] Bracha, V., Wu, J. Z., Cartwright, S., and Bloedel, J. R. (1991). Selective involvement of the spinal trigeminal nucleus in the conditioned nictitating membrane reflex of the rabbit. *Brain research*, 556(2):317–20.

[Braitenberg, 1986] Braitenberg, V. (1986). *Vehicles: Experiments in synthetic psychology*. MIT press.

[Braitenberg, 1990] Braitenberg, V. (1990). Reading the structure of brains. *Network: Computation in Neural Systems*, 1(1):1–11.

[Braitenberg and Atwood, 1958] Braitenberg, V. and Atwood, R. P. (1958). Morphological observations on the cerebellar cortex. *The Journal of Comparative Neurology*, 109(1):1—-27.

[Brandi et al., 2013] Brandi, S., Herreros, I., Sánchez-Fibla, M., and Verschure, P. F. M. J. (2013). Learning of Motor Sequences Based on a Computational Model of the Cerebellum. *Biomimetic and Biohybrid Systems*.

[Brooks, 1985] Brooks, R. A. (1985). A Robust Layered Control System For a Mobile Robot.

[Brooks, 1991] Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence*, 47(1-3):139–159.

[Brown et al., 1977] Brown, J. T., Chan-Palay, V., and Palay, S. L. (1977). A study of afferent input to the inferior olivary complex in the rat by retrograde axonal transport of horseradish peroxidase. *The Journal of comparative neurology*, 176(1):1–22.

[Brun et al., 2008] Brun, V. H., Solstad, T., Kjelstrup, K. B., Fyhn, M., Witter, M. P., Moser, E. I., and Moser, M.-B. (2008). Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. *Hippocampus*, 18(12):1200–12.

[Burgess et al., 2002] Burgess, N., Maguire, E. A., and O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, 35(4):625–41.

[Buzsáki and Draguhn, 2004] Buzsáki, G. and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *science*, 304(5679):1926–1929.

[Buzsáki and Moser, 2013] Buzsáki, G. and Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature neuroscience*, 16(2):130–8.

[Cabanac and Swiergiel, 1989] Cabanac, M. and Swiergiel, A. H. (1989). Rats eating and hoarding as a function of body weight and cost of foraging. *The American journal of physiology*, 257(4 Pt 2):R952–7.

[Caligiore et al., 2017] Caligiore, D., Pezzulo, G., Baldassarre, G., Bostan, A. C., Strick, P. L., Doya, K., Helmich, R. C., Dirkx, M., Houk, J., Jörntell, H., et al. (2017). Consensus paper: towards a systems-level view of cerebellar function: the interplay between cerebellum, basal ganglia, and cortex. *The Cerebellum*, 16(1):203–229.

[Cavanagh and Frank, 2014] Cavanagh, J. F. and Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in cognitive sciences*, 18(8):414–421.

[Cavanagh et al., 2010] Cavanagh, J. F., Frank, M. J., Klein, T. J., and Allen, J. J. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *Neuroimage*, 49(4):3198–3209.

[Cavanagh et al., 2011] Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., and Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature neuroscience*, 14(11):1462.

[Cerminara et al., 2015] Cerminara, N. L., Lang, E. J., Sillitoe, R. V., and Apps, R. (2015). Redefining the cerebellar cortex as an assembly of non-uniform purkinje cell microcircuits. *Nature Reviews Neuroscience*, 16(2):79.

[Charnov, 1976] Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, 9(2):129–136.

[Chater et al., 2006] Chater, N., Tenenbaum, J. B., and Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations.

[Chen and Wise, 1997] Chen, L. L. and Wise, S. P. (1997). Conditional oculomotor learning: population vectors in the supplementary eye field. *J Neurophysiol*, 78(2):1166–1169.

[Chen et al., 2010] Chen, X., Scangos, K. W., and Stuphorn, V. (2010). Supplementary motor area exerts proactive and reactive control of arm movements. *Journal of Neuroscience*, 30(44):14657–14675.

[Christian and Thompson,] Christian, K. M. and Thompson, R. F. Neural substrates of eyeblink conditioning: acquisition and retention. *Learning & memory (Cold Spring Harbor, N.Y.)*, 10(6):427–55.

[Clark and Squire, 1998] Clark, R. E. and Squire, L. R. (1998). Classical conditioning and brain systems: the role of awareness. *Science (New York, N.Y.)*, 280(5360):77–81.

[Clayton and Dickinson, 1998] Clayton, N. S. and Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, 395(6699):272–4.

[Clopath et al., 2014] Clopath, C., Badura, A., De Zeeuw, C. I., and Brunel, N. (2014). A cerebellar learning model of vestibulo-ocular reflex adaptation in wild-type and mutant mice. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 34(21):7203–15.

[Cohen, 2014a] Cohen, M. X. (2014a). A neural microcircuit for cognitive conflict detection and signaling. *Trends in neurosciences*, 37(9):480–490.

[Cohen, 2014b] Cohen, M. X. (2014b). A neural microcircuit for cognitive conflict detection and signaling. *Trends in neurosciences*, 37(9):480–490.

[Cohen and Cavanagh, 2011] Cohen, M. X. and Cavanagh, J. F. (2011). Single-trial regression elucidates the role of prefrontal theta oscillations in response conflict. *Frontiers in psychology*, 2:30.

[Corbit and Balleine, 2000] Corbit, L. H. and Balleine, B. W. (2000). The Role of the Hippocampus in Instrumental Conditioning. *J. Neurosci.*, 20(11):4233–4239.

[Cordeschi, 2002] Cordeschi, R. (2002). *The Discovery of the Artificial: Behavior, Mind and Machines Before and Beyond Cybernetics*. Springer Science & Business Media.

[Cordo and Nashner, 1982] Cordo, P. and Nashner, L. M. (1982). Properties of postural adjustments associated with rapid arm movements. *Journal of neurophysiology*, 47(2):287–302.

[Crevecoeur and Scott, 2013] Crevecoeur, F. and Scott, S. H. (2013). Priors engaged in long-latency responses to mechanical perturbations suggest a rapid update in state estimation. *PLoS computational biology*, 9(8):e1003177.

[Criscimagna-Hemminger et al., 2003] Criscimagna-Hemminger, S. E., Donchin, O., Gazzaniga, M. S., and Shadmehr, R. (2003). Learned dynamics of reaching movements generalize from dominant to nondominant arm. *Journal of neurophysiology*, 89(1):168–176.

[Crone et al., 2006] Crone, E. A., Wendelken, C., Donohue, S. E., and Bunge, S. A. (2006). Neural evidence for dissociable components of task-switching. *Cerebral Cortex*, 16(4):475–486.

[Damasio and Sutherland, 1994] Damasio, A. R. and Sutherland, S. (1994). Descartes' error: Emotion, reason and the human brain. *Nature*, 372(6503):287–287.

[D'Ardenne et al., 2012] D'Ardenne, K., Eshel, N., Luka, J., Lenartowicz, A., Nystrom, L. E., and Cohen, J. D. (2012). Role of prefrontal cortex and the midbrain dopamine system in working memory updating. *Proceedings of the National Academy of Sciences*, 109(49):19900–19909.

[Dasgupta et al., 2014] Dasgupta, S., Wörgötter, F., and Manoonpong, P. (2014). Neuromodulatory adaptive combination of correlation-based learning in cerebellum and reward-based learning in basal ganglia for goal-directed behavior control. *Frontiers in neural circuits*, 8:126.

[Daw, 2012] Daw, N. D. (2012). Model-based reinforcement learning as cognitive search: neurocomputational theories. *Cognitive search: Evolution, algorithms and the brain*, pages 195–208.

[Dayan and Berridge, 2014] Dayan, P. and Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation. *Cognitive, affective & behavioral neuroscience*, 14(2):473–92.

[de Almeida et al., 2009] de Almeida, L., Idiart, M., and Lisman, J. E. (2009). The input-output transformation of the hippocampal granule cells: from grid cells to place fields. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(23):7504–12.

[Dean and Porrill, 2011] Dean, P. and Porrill, J. (2011). Evaluating the adaptive-filter model of the cerebellum. *The Journal of physiology*, 589(14):3459–3470.

[Dean et al., 2010a] Dean, P., Porrill, J., Ekerot, C.-F., and Jörntell, H. (2010a). The cerebellar microcircuit as an adaptive filter: experimental and computational evidence. *Nature Reviews Neuroscience*, 11(1):30.

[Dean et al., 2010b] Dean, P., Porrill, J., Ekerot, C.-F., and Jörntell, H. (2010b). The cerebellar microcircuit as an adaptive filter: experimental and computational evidence. *Nature reviews. Neuroscience*, 11(1):30–43.

[Deecke and Kornhuber, 1978] Deecke, L. and Kornhuber, H. H. (1978). An electrical sign of participation of the mesial 'supplementary' motor cortex in human voluntary finger movement. *Brain Research*, 159(2):473–476.

[Dehaene et al., 2014] Dehaene, S., Charles, L., King, J.-R., and Marti, S. (2014). Toward a computational theory of conscious processing. *Current opinion in neurobiology*, 25:76–84.

[Delorme et al., 2007] Delorme, A., Westerfield, M., and Makeig, S. (2007). Medial prefrontal theta bursts precede rapid motor responses during visual selective attention. *Journal of Neuroscience*, 27(44):11949–11959.

[Deluca et al., 2014] Deluca, C., Golzar, A., Santandrea, E., Lo Gerfo, E., Eštočinová, J., Moretto, G., Fiaschi, A., Panzeri, M., Mariotti, C., Tinazzi, M., and Chelazzi, L. (2014). The cerebellum and visual perceptual learning: Evidence from a motion extrapolation task. *Cortex; a journal devoted to the study of the nervous system and behavior*, 58:52–71.

[Diedrichsen et al., 2005] Diedrichsen, J., Verstynen, T., Lehman, S. L., and Ivry, R. B. (2005). Cerebellar involvement in anticipating the consequences of self-produced actions during bimanual movements. *Journal of Neurophysiology*, 93(2):801–812.

[Dissanayake et al., 2001] Dissanayake, M., Newman, P., Clark, S., Durrant-Whyte, H., and Csorba, M. (2001). A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotics and Automation*, 17(3):229–241.

[Ditterich, 2006] Ditterich, J. (2006). Stochastic models of decisions about motion direction: behavior and physiology. *Neural networks : the official journal of the International Neural Network Society*, 19(8):981–1012.

[Domjan, 2004] Domjan, M. (2004). Pavlovian Conditioning: A Functional Perspective.

[Donegan et al., 1989] Donegan, N. H., Gluck, M. A., and Thompson, R. F. (1989). Integrating behavioral and biological models of classical conditioning. In *Psychology of learning and motivation*, volume 23, pages 109–156. Elsevier.

[Doya, 1999] Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural networks : the official journal of the International Neural Network Society*, 12(7-8):961–974.

[Doya, 2000] Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current opinion in neurobiology*, 10(6):732–739.

[Duff et al., 2011] Duff, A., Fibla, M. S., and Verschure, P. F. M. J. (2011). A biologically based model for the integration of sensory-motor contingencies in rules and plans: a prefrontal cortex based extension of the Distributed Adaptive Control architecture. *Brain research bulletin*, 85(5):289–304.

[Duff and Verschure, 2010] Duff, A. and Verschure, P. F. (2010). Unifying perceptual and behavioral learning with a correlative subspace learning rule. *Neurocomputing*, 73(10-12):1818–1830.

[Dugatkin, 2008] Dugatkin, L. A. (2008). Principles of Animal Behavior (Second Edition).

[Dum and Strick, 2003] Dum, R. P. and Strick, P. L. (2003). An unfolded map of the cerebellar dentate nucleus and its projections to the cerebral cortex. *Journal of neurophysiology*, 89:634–639.

[Eccles, 1982] Eccles, J. C. (1982). The initiation of voluntary movements by the supplementary motor area. *Archives of Psychiatry and Neurological Sciences*, 231:423–441.

[Eccles et al.,] Eccles, J. C., Ito, M., and Szentagothai, J. The cerebellum as a neuronal machine. 1967. *Google Scholar*, pages 1–335.

[Eccles et al., 1967] Eccles, J. C., Ito, M., and Szentágothai, J. (1967). The mossy fiber input into the cerebellar cortex and its inhibitory control by golgi cells. In *The cerebellum as a neuronal machine*, pages 116–155. Springer.

[Edelman, 2007] Edelman, G. M. (2007). Learning in and from brain-based devices. *Science (New York, N.Y.)*, 318(5853):1103–5.

[Eliasmith, 2005] Eliasmith, C. (2005). A unified approach to building and controlling spiking attractor networks. *Neural computation*, 17(6):1276–314.

[Eliasmith et al., 2012] Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., Tang, C., and Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science (New York, N.Y.)*, 338(6111):1202–5.

[Engel et al., 2001] Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience*, 2(10):704.

[Enochson and Goodman, 1965] Enochson, L. D. and Goodman, N. R. (1965). Gaussian approximations to the distribution of sample coherence. Technical report, MEASUREMENT ANALYSIS CORP LOS ANGELES CA.

[Erdem and Hasselmo, 2014] Erdem, U. M. and Hasselmo, M. E. (2014). A biologically inspired hierarchical goal directed navigation model. *Journal of physiology, Paris*, 108(1):28–37.

[Fagg, 2002] Fagg, A. H. (2002). A Computational Model of Muscle Recruitment for Wrist Movements. *Journal of Neurophysiology*, 88(6):3348–3358.

[Fantino and Abarca, 2010] Fantino, E. and Abarca, N. (2010). Choice, optimal foraging, and the delay-reduction hypothesis. *Behavioral and Brain Sciences*, 8(02):315.

[Fedorov et al., 2012] Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al. (2012). 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9):1323–1341.

[Fibla et al., 2010] Fibla, M. S., Bernardet, U., and Verschure, P. F. M. J. (2010). Allostatic control for robot behaviour regulation: An extension to path planning. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1935–1942. IEEE.

[Fischl, 2012] Fischl, B. (2012). Freesurfer. *Neuroimage*, 62(2):774–781.

[Flanagan et al., 2001] Flanagan, J. R., King, S., Wolpert, D. M., and Johansson, R. S. (2001). Sensorimotor prediction and memory in object manipulation. *Canadian Journal of Experimental Psychology*, 55.

[Fodor, 1985] Fodor, J. A. (1985). Precis of the modularity of mind. *Behavioral and brain sciences*, 8(1):1–5.

[Friend and Kravitz, 2014] Friend, D. M. and Kravitz, A. V. (2014). Working together: basal ganglia pathways in action selection. *Trends in neurosciences*, 37(6):301–303.

[Fries, 2015] Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron*, 88(1):220–235.

[Friston, 2010] Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127.

[Friston, 2011a] Friston, K. (2011a). What is optimal about motor control? *Neuron*, 72(3):488–498.

[Friston, 2011b] Friston, K. (2011b). What Is Optimal about Motor Control? *Neuron*, 72(3):488–498.

[Friston et al., 2015] Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive neuroscience*, 6(4):1–28.

[Friston et al., 2010] Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biological cybernetics*, 102(3):227–260.

[Fujita, 1982] Fujita, M. (1982). Adaptive filter model of the cerebellum. *Biological cybernetics*, 45(3):195–206.

[Fukuda et al., 1972] Fukuda, J., Highstein, S. M., and Ito, M. (1972). Cerebellar inhibitory control of the vestibulo-ocular reflex investigated in rabbit IIIrd nucleus. *Experimental Brain Research*, 14(5):511–526.

[Fuster, 2015] Fuster, J. (2015). *The prefrontal cortex*. Academic Press.

[Gage et al., 2004] Gage, W. H., Winter, D. A., Frank, J. S., and Adkin, A. L. (2004). Kinematic and kinetic validity of the inverted pendulum model in quiet standing. *Gait & posture*, 19(2):124–32.

[Gao et al., 1996] Gao, J.-H., Parsons, L. M., Bower, J. M., Xiong, J., Li, J., and Fox, P. T. (1996). Cerebellum Implicated in Sensory Acquisition and Discrimination Rather Than Motor Control. *Science*, 272(5261):545–547.

[Gold and Shadlen, 2007] Gold, J. I. and Shadlen, M. N. (2007). The neural basis of decision making. *Annual review of neuroscience*, 30:535–74.

[Goldberg, 1985] Goldberg, G. (1985). Supplementary motor area structure and function: review and hypotheses. *Behav.Brain Sci.*, 8(1985):567–616.

[Gomi and Kawato, 1992] Gomi, H. and Kawato, M. (1992). Adaptive feedback control models of the vestibulocerebellum and spinocerebellum. *Biol.Cybern.*, 68(2):105–114.

[Gomi and Kawato, 1993] Gomi, H. and Kawato, M. (1993). Neural network control for a closed-loop system using feedback-error-learning. *Neural Networks*, 6(7):933–946.

[Goodale et al., 1975] Goodale, M., Milner, A., and Rose, J. (1975). Susceptibility to startle during ongoing behaviour following collicular lesions in the rat. *Neuroscience Letters*, 1(6):333–337.

[Gormezano, 1972] Gormezano, I. (1972). Investigations of defense and reward conditioning in the rabbit. *Classical conditioning II: Current research and theory*, pages 151–181.

[Gormezano et al.,] Gormezano, I., Prokasy, W. F., and Thompson, R. F. *Classical conditioning (3rd ed.).*

[Gould et al., 2010] Gould, K. L., Kelly, D. M., and Kamil, A. C. (2010). What scatter-hoarding animals have taught us about small-scale navigation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1542):901–14.

[Grace et al., 2007] Grace, A. A., Floresco, S. B., Goto, Y., and Lodge, D. J. (2007). Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. *Trends in neurosciences*, 30(5):220–227.

[Guanella et al., 2007] Guanella, A., Kiper, D., and Verschure, P. (2007). A model of grid cells based on a twisted torus topology. *International journal of neural systems*, 17(4):231–40.

[Guell et al., 2018] Guell, X., Schmahmann, J. D., Gabrieli, J. D., and Ghosh, S. S. (2018). Functional Gradients of The Cerebellum: A Fundamental Movement-to-thought Principle. pages 1–43.

[Guitart-Masip et al., 2013] Guitart-Masip, M., Barnes, G. R., Horner, A., Bauer, M., Dolan, R. J., and Duzel, E. (2013). Synchronization of medial temporal lobe and prefrontal rhythms in human decision making. *Journal of Neuroscience*, 33(2):442–451.

[Hafting et al., 2005] Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., and Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–6.

[Hassabis et al., 2017] Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-Inspired Artificial Intelligence. *Neuron*, 95(2):245–258.

[Helfrich and Knight, 2016] Helfrich, R. F. and Knight, R. T. (2016). Oscillatory Dynamics of Prefrontal Cognitive Control. *Trends in Cognitive Sciences*, 20(12):916–930.

[Herreros et al., 2016] Herreros, I., Arsiwalla, X., and Verschure, P. (2016). A forward model at purkinje cell synapses facilitates cerebellar anticipatory control. In *Advances in Neural Information Processing Systems*, pages 3828–3836.

[Herreros et al., 2013a] Herreros, I., Maffei, G., Brandi, S., Sánchez-Fibla, M., and Verschure, P. F. (2013a). Speed generalization capabilities of a cerebellar model on a rapid navigation task. In *Intelligent*

Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on, pages 363–368. IEEE.

[Herreros et al., 2013b] Herreros, I., Maffei, G., Brandi, S., Sanchez-Fibla, M., and Verschure, P. F. M. J. (2013b). Speed generalization capabilities of a cerebellar model on a rapid navigation task. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 363–368. IEEE.

[Herreros and Verschure, 2013a] Herreros, I. and Verschure, P. F. (2013a). Nucleo-olivary inhibition balances the interaction between the reactive and adaptive layers in motor control. *Neural Networks*, 47:64–71.

[Herreros and Verschure, 2013b] Herreros, I. and Verschure, P. F. (2013b). Nucleo-olivary inhibition balances the interaction between the reactive and adaptive layers in motor control. *Neural Networks*, 47:64–71.

[Herreros and Verschure, 2015] Herreros, I. and Verschure, P. F. (2015). About the goal of a goals’ goal theory. *Cognitive neuroscience*, 6(4):218–9.

[Hesslow, 1994] Hesslow, G. (1994). Inhibition of classically conditioned eyeblink responses by stimulation of the cerebellar cortex in the decerebrate cat. *J. Physiol.*, 476(2):245–256.

[Hikosaka and Isoda, 2010a] Hikosaka, O. and Isoda, M. (2010a). Switching from automatic to controlled behavior: cortico-basal ganglia mechanisms. *Trends in Cognitive Sciences*, 14(4):154–161.

[Hikosaka and Isoda, 2010b] Hikosaka, O. and Isoda, M. (2010b). Switching from automatic to controlled behavior: cortico-basal ganglia mechanisms. *Trends in cognitive sciences*, 14(4):154–161.

[Hipp et al., 2011] Hipp, J. F., Engel, A. K., and Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron*, 69(2):387–396.

[HofstoÈtter et al., 2002] HofstoÈtter, C., Mintz, M., and Verschure, P. F. (2002). The cerebellum in action: a simulation and robotics study. *European Journal of Neuroscience*, 16(7):1361–1376.

[Horak et al., 1994] Horak, F. B., Diener, H. C., and Neurological, R. S. D. (1994). Cerebellar control of postural scaling and central set in stance Cerebellar Control of Postural Scaling and Central Set in Stance. *Journal of Neurophysiology*, 72(2):479–493.

[Hoshi et al., 2005] Hoshi, E., Tremblay, L., Féger, J., Carras, P. L., and Strick, P. L. (2005). The cerebellum communicates with the basal ganglia. *Nature neuroscience*, 8(11):1491.

[Hull,] Hull, C. L. *Essentials of behavior.*

[Hyafil et al., 2015] Hyafil, A., Giraud, A. L., Fontolan, L., and Gutkin, B. (2015). Neural Cross-Frequency Coupling: Connecting Architectures, Mechanisms, and Functions. *Trends in Neurosciences*, 38(11):725–740.

[Inderbitzin et al., 2010] Inderbitzin, M., Herreros-Alonso, I., and Verschure, P. F. (2010). An integrated computational model of the two phase theory of classical conditioning. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

[Iodice et al., 2015] Iodice, P., Cesinaro, S., Romani, G. L., and Pezzulo, G. (2015). More gain less pain: balance control learning shifts the activation patterns of leg and neck muscles and increases muscular parsimony. *Experimental brain research*, 233(7):2103–2114.

[Isoda and Hikosaka, 2007] Isoda, M. and Hikosaka, O. (2007). Switching from automatic to controlled action by monkey medial frontal cortex. *Nature neuroscience*, 10(2):240.

[Isoda and Tanji, 2004] Isoda, M. and Tanji, J. (2004). Participation of the primate presupplementary motor area in sequencing multiple saccades. *Journal of neurophysiology*, 92(1):653–659.

[Ito et al., 2015] Ito, H. T., Zhang, S.-J., Witter, M. P., Moser, E. I., and Moser, M.-B. (2015). A prefrontal-thalamo-hippocampal circuit for goal-directed spatial navigation. *Nature*, 522(7554):50–5.

[Ito, 1987] Ito, M. (1987). Signal processing in cerebellar purkinje cells. *Physiologia Bohemoslovaca*, 36(3):203–216.

[Ito, 2006] Ito, M. (2006). Cerebellar circuitry as a neuronal machine. *Progress in neurobiology*, 78(3-5):272–303.

[Ito, 2008] Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature reviews. Neuroscience*, 9(4):304–13.

[Ito et al., 1982] Ito, M., Sakurai, M., and Tongroach, P. (1982). Climbing fibre induced depression of both mossy fibre responsiveness and glutamate sensitivity of cerebellar purkinje cells. *The Journal of Physiology*, 324(1):113–134.

[Jackendoff, 1987] Jackendoff, R. (1987). *Consciousness and the computational mind.* The MIT Press.

[Jarvis and Mitra, 2001] Jarvis, M. and Mitra, P. (2001). Sampling properties of the spectrum and coherency of sequences of action potentials. *Neural Computation*, 13(4):717–749.

[Jeneson and Squire, 2012] Jeneson, A. and Squire, L. (2012). Working memory, long-term memory, and medial temporal lobe function. *Learning & Memory*, 19(1):15–25.

[Jerison, 1971] Jerison, H. (1971). More on why birds and mammals have big brains. *The American Naturalist*, 105(942):185–189.

[Jerison, 2012] Jerison, H. (2012). *Evolution of the brain and intelligence*. Elsevier.

[Jerison and Barlow, 1985] Jerison, H. J. and Barlow, H. B. (1985). Animal Intelligence as Encephalization [and Discussion]. *Philosophical*

Transactions of the Royal Society B: Biological Sciences, 308(1135):21–35.

[Jirenhed et al., 2007] Jirenhed, D.-A., Bengtsson, F., and Hesslow, G. (2007). Acquisition, extinction, and reacquisition of a cerebellar cortical memory trace. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(10):2493–502.

[Johansson et al., 2014] Johansson, F., Jirenhed, D.-A., Rasmussen, A., Zucca, R., and Hesslow, G. (2014). Memory trace and timing mechanism localized to cerebellar purkinje cells. *Proceedings of the National Academy of Sciences*, 111(41):14930–14934.

[Johnson and Redish, 2007] Johnson, A. and Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(45):12176–89.

[Jurado-Parras et al., 2013] Jurado-Parras, M. T., Sánchez-Campusano, R., Castellanos, N. P., Del-Pozo, F., Gruart, A., and Delgado-García, J. M. (2013). Differential contribution of hippocampal circuits to appetitive and consummatory behaviors during operant conditioning of behaving mice. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33(6):2293–304.

[Kaas, 2013] Kaas, J. H. (2013). The evolution of brains from early mammals to humans. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(1):33–45.

[Kahneman, 2011] Kahneman, D. (2011). *Thinking, fast and slow*.

[Kawamura et al., 1982] Kawamura, S., Hattori, S., Higo, S., and Matsuyama, T. (1982). The cerebellar projections to the superior colliculus and pretectum in the cat: an autoradiographic and horseradish peroxidase study. *Neuroscience*, 7(7):1673–89.

[Kawato, 1999] Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9(6):718–727.

[Kawato et al., 1987] Kawato, M., Furukawa, K., and Suzuki, R. (1987). A hierarchical neural-network model for control and learning of voluntary movement. *Biological cybernetics*, 57(3):169–185.

[Kawato and Gomi, 1992] Kawato, M. and Gomi, H. (1992). A computational model of four regions of the cerebellum based on feedback-error learning. *Biological cybernetics*, 68(2):95–103.

[Keele et al., 2003] Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., and Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological review*, 110(2):316.

[Keen-Rhinehart et al., 2010] Keen-Rhinehart, E., Dailey, M. J., and Bartness, T. (2010). Physiological mechanisms for food-hoarding motivation in animals. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1542):961–75.

[Kelly and Strick, 2003] Kelly, R. M. and Strick, P. L. (2003). Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *Journal of neuroscience*, 23(23):8432–8444.

[Kettner et al., 1997] Kettner, R. E., Mahamud, S., Leung, H. C., Sitkoff, N., Houk, J. C., Peterson, B. W., and Barto, a. G. (1997). Prediction of complex two-dimensional trajectories by a cerebellar model of smooth pursuit eye movement. *Journal of neurophysiology*, 77:2115–2130.

[Kim et al., 2013] Kim, E. J., Horovitz, O., Pellman, B. A., Tan, L. M., Li, Q., Richter-Levin, G., and Kim, J. J. (2013). Dorsal periaqueductal gray-amygdala pathway conveys both innate and learned fear responses in rats. *Proceedings of the National Academy of Sciences of the United States of America*, 110(36):14795–800.

[Knierim et al., 2014] Knierim, J. J., Neunuebel, J. P., and Deshmukh, S. S. (2014). Functional correlates of the lateral and medial entorhinal cortex: objects, path integration and local-global reference frames. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1635):20130369.

[Kobayashi and Schultz, 2008] Kobayashi, S. and Schultz, W. (2008). Influence of reward delays on responses of dopamine neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28(31):7837–46.

[Koechlin, 2016] Koechlin, E. (2016). Prefrontal executive function and adaptive behavior in complex environments. *Current Opinion in Neurobiology*, 37:1–6.

[Koechlin et al., 2003] Koechlin, E., Ody, C., and Kouneiher, F. (2003). The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science*, 302(5648):1181–1185.

[Kolb et al., 2004] Kolb, F., Lachauer, S., Maschke, M., and Timmann, D. (2004). Classically conditioned postural reflex in cerebellar patients. *Experimental brain research*, 158(2):163–179.

[Konorski and Miller, 1937] Konorski, J. and Miller, S. (1937). On Two Types of Conditioned Reflex. *The Journal of General Psychology*, 16(1):264–272.

[Krichmar et al., 2005] Krichmar, J. L., Nitz, D. A., Gally, J. A., and Edelman, G. M. (2005). Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task. *Proceedings of the National Academy of Sciences*, 102(6):2111–2116.

[Lachaux et al., 1999] Lachaux, J.-P., Rodriguez, E., Martinerie, J., Varela, F. J., et al. (1999). Measuring phase synchrony in brain signals. *Human brain mapping*, 8(4):194–208.

[Lacquaniti and Maioli, 1989] Lacquaniti, F. and Maioli, C. (1989). The role of preparation in tuning anticipatory and reflex responses during catching. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 9(1):134–48.

[Laird, 1987] Laird, J. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence*, 33(1):1–64.

[Laird, 2012] Laird, J. E. (2012). *The Soar cognitive architecture*. MIT press.

[Lallee et al., 2013] Lallee, S., Hamann, K., Steinwender, J., Warneken, F., Martienz, U., Barron-Gonzales, H., Pattacini, U., Gori, I., Petit, M., Metta, G., Verschure, P., and Ford Dominey, P. (2013). Cooperative human robot interaction systems: IV. Communication of shared plans with Naïve humans using gaze and speech. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 129–136. IEEE.

[Lang and Bastian, 1999] Lang, C. and Bastian, A. (1999). Cerebellar subjects show impaired adaptation of anticipatory EMG during catching. *Journal of Neurophysiology*, pages 2108–2119.

[Latash, 2008] Latash, M. L. (2008). *Neurophysiological basis of movement*. Human Kinetics.

[Lea and Tarpy, 1986] Lea, S. E. and Tarpy, R. M. (1986). Hamsters' demand for food to eat and hoard as a function of deprivation and cost. *Animal Behaviour*, 34(6):1759–1768.

[LeDoux, 2012] LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, 73(4):653–76.

[Li and Aruin, 2007] Li, X. and Aruin, A. S. (2007). The effect of short-term changes in the body mass on anticipatory postural adjustments. *Experimental brain research*, 181(2):333–46.

[Lisman, 2005] Lisman, J. (2005). The theta/gamma discrete phase code occuring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus*, 15(7):913–22.

[Lisman and Jensen, 2013] Lisman, J. E. and Jensen, O. (2013). The theta-gamma neural code. *Neuron*, 77(6):1002–1016.

[Lockie and Andrews, 2013] Lockie, S. H. and Andrews, Z. B. (2013). The hormonal signature of energy deficit: Increasing the value of food reward. *Molecular metabolism*, 2(4):329–36.

[Lopour et al., 2013] Lopour, B. A., Tavassoli, A., Fried, I., and Ringach, D. L. (2013). Coding of information in the phase of local field potentials within human medial temporal lobe. *Neuron*, 79(3):594–606.

[Lu et al., 2012] Lu, X., Miyachi, S., and Takada, M. (2012). Anatomical evidence for the involvement of medial cerebellar output from the interpositus nuclei in cognitive functions. *Proceedings of the National Academy of Sciences*, 109(46):18980–18984.

[Luo et al., 2011] Luo, A. H., Tahsili-Fahadan, P., Wise, R. A., Lupica, C. R., and Aston-Jones, G. (2011). Linking context with reward: a functional circuit from hippocampal CA3 to ventral tegmental area. *Science (New York, N.Y.)*, 333(6040):353–7.

[MacKinnon et al., 2007] MacKinnon, C. D., Bissig, D., Chiusano, J., Miller, E., Rudnick, L., Jager, C., Zhang, Y., Mille, M.-L., and Rogers, M. W. (2007). Preparation of anticipatory postural adjustments prior to stepping. *Journal of Neurophysiology*, 97(6):4368–4379.

[Mackintosh, 1983] Mackintosh, N. J. (1983). *Conditioning and associative learning*.

[Maffei et al., 2013] Maffei, G., Herreros, I., Sánchez-Fibla, M., and Verschure, P. F. (2013). Acquisition of anticipatory postural adjustment through cerebellar learning in a mobile robot. In *Conference on Biomimetic and Biohybrid Systems*, pages 399–401. Springer.

[Maffei et al., 2014] Maffei, G., Sanchez-Fibla, M., Herreros, I., and Verschure, P. F. (2014). The role of a cerebellum-driven perceptual prediction within a robotic postural task. In *International Conference on Simulation of Adaptive Behavior*, pages 76–87. Springer.

[Marcos et al., 2012] Marcos, E., Duff, A., Sanchez-Fibla, M., and Verschure, P. F. M. J. (2012). *Generalization of Integrator Models to Foraging: A Robot Study Using the DAC9 Model*, volume 7375 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, biomimetic edition.

[Marcos, Encarni Verschure, 2013] Marcos, Encarni Verschure, P. F. M. J. (2013). *Computational and Robotic Models of the Hierarchical Organization of Behavior*. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Maris and Oostenveld, 2007] Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of eeg-and meg-data. *Journal of neuroscience methods*, 164(1):177–190.

[Maris et al., 2007] Maris, E., Schoffelen, J.-M., and Fries, P. (2007). Nonparametric statistical testing of coherence differences. *Journal of neuroscience methods*, 163(1):161–175.

[Marr, 1969] Marr, D. (1969). A THEORY OF CEREBELLAR CORTEX. *J. Physiol*, 202:437–470.

[Massion, 1992] Massion, J. (1992). MOVEMENT , POSTURE AND EQUILIBRIUM : INTERACTION AND COORDINATION Control of Feedforward gain and gate control Perturbation. *Progress in Neurobiology*, 38.

[Massion, 1994] Massion, J. (1994). Postural control system. *Current Opinion in Neurobiology*, 4(6):877–887.

[Mathews et al., 2015] Mathews, Z., Cetnarski, R., and Verschure, P. F. M. J. (2015). Visual anticipation biases conscious decision making but not bottom-up visual processing. *Frontiers in Psychology*, 5:1443.

[Mathis et al., 2017] Mathis, M. W., Mathis, A., and Uchida, N. (2017). Somatosensory cortex plays an essential role in forelimb motor adaptation in mice. *Neuron*, 93(6):1493–1503.

[May, 2006] May, P. J. (2006). The mammalian superior colliculus: laminar structure and connections. *Progress in brain research*, 151:321–78.

[McIntyre et al., 1995] McIntyre, J., Gurfinkel, E., Lipshits, M., Droulez, J., and Gurfinkel, V. (1995). Measurements of human force control during a constrained arm motion using a force-actuated joystick. *Journal of neurophysiology*, 73(3):1201–1222.

[McKinstry et al., 2006] McKinstry, J. L., Edelman, G. M., and Krichmar, J. L. (2006). A cerebellar model for predictive motor control tested in a brain-based device. *Proceedings of the National Academy of Sciences of the United States of America*, 103(9):3387–92.

[Medin et al., 1996] Medin, D. L., Shanks, D. R., and Holyoak, K. J. (1996). *Causal Learning: Advances in Research and Theory*, volume 34. Academic Press.

[Medina et al., 2002] Medina, J. F., Repa, J. C., Mauk, M. D., and LeDoux, J. E. (2002). Parallels between cerebellum- and amygdala-dependent conditioning. *Nature reviews. Neuroscience*, 3(2):122–31.

[Meredith and Stein, 1983] Meredith, M. and Stein, B. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, 221(4608):389–391.

[Miall, 1998] Miall, R. (1998). The cerebellum, predictive control and motor coordination. *Sensory guidance of movement*, 218:272–290.

[Miall et al., 1993] Miall, R., Weir, D. J., Wolpert, D. M., and Stein, J. (1993). Is the cerebellum a smith predictor? *Journal of motor behavior*, 25(3):203–216.

[Miall and Wolpert, 1996] Miall, R. and Wolpert, D. (1996). Forward Models for Physiological Motor Control. *Neural Networks*, 9(8):1265–1279.

[Middleton and Strick, 1998] Middleton, F. A. and Strick, P. L. (1998). Cerebellar output: Motor and cognitive channels. *Trends in Cognitive Sciences*, 2(9):348–355.

[Miles and Lisberger, 1981] Miles, F. A. and Lisberger, S. G. (1981). Plasticity in the vestibulo-ocular reflex: a new hypothesis. *Annual review of neuroscience*, 4:273–99.

[Milford and Schulz, 2014] Milford, M. and Schulz, R. (2014). Principles of goal-directed spatial robot navigation in biomimetic models. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1655):20130484–.

[Miłkowski, 2013] Miłkowski, M. (2013). *Explaining the computational mind*. Mit Press.

[Miller and Cohen, 2001] Miller, E. K. and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24:167–202.

[Mitra and Pesaran, 1999] Mitra, P. P. and Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophysical journal*, 76(2):691–708.

[Modayil and Sutton, 2014] Modayil, J. and Sutton, R. (2014). Prediction Driven Behavior: Learning Predictions that Drive Fixed Responses. In *AAAI Workshop on AI and Robotics*.

[Mohapatra et al., 2012] Mohapatra, S., Krishnan, V., and Aruin, A. S. (2012). Postural control in response to an external perturbation: effect of altered proprioceptive information. *Experimental brain research*, 217(2):197–208.

[Moisello et al., 2009] Moisello, C., Crupi, D., Tunik, E., Quartarone, A., Bove, M., Tononi, G., and Ghilardi, M. F. (2009). The serial reaction time task revisited: a study on motor sequence learning with an arm-reaching task. *Experimental brain research*, 194(1):143–155.

[Montague et al., 1995] Montague, P. R., Dayan, P., Person, C., and Sejnowski, T. J. (1995). Bee foraging in uncertain environments using predictive hebbian learning. *Nature*, 377(6551):725–8.

[Monzée and Smith, 2004] Monzée, J. and Smith, A. M. (2004). Responses of cerebellar interpositus neurons to predictable perturbations applied to an object held in a precision grip. *Journal of neurophysiology*, 91(3):1230–1239.

[Morton et al., 2001] Morton, S. M., Lang, C. E., and Bastian, A. J. (2001). Inter-and intra-limb generalization of adaptation during catching. *Experimental Brain Research*, 141(4):438–445.

[Mushiake et al., 1990] Mushiake, H., Inase, M., and Tanji, J. (1990). Selective coding of motor sequence in the supplementary motor area of the monkey cerebral cortex. *Experimental Brain Research*, 82(1):208–210.

[Nachev et al., 2008] Nachev, P., Kennard, C., and Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience*, 9(11):856–869.

[Nachev et al., 2005] Nachev, P., Rees, G., Parton, A., Kennard, C., and Husain, M. (2005). Volition and conflict in human medial frontal cortex. *Current Biology*, 15(2):122–128.

[Nakamura et al., 1998] Nakamura, K., Sakai, K., and Hikosaka, O. (1998). Neuronal activity in medial frontal cortex during learning of sequential procedures. *J Neurophysiol*, 80(5):2671–2687.

[Narayanan et al., 2013] Narayanan, N. S., Cavanagh, J. F., Frank, M. J., and Laubach, M. (2013). Common medial frontal mechanisms of adaptive control in humans and rodents. *Nature neuroscience*, 16(12):1888.

[Newell, 1994] Newell, A. (1994). *Unified Theories of Cognition*. Harvard University Press.

[Newell et al., 1958] Newell, A., Shaw, J. C., and Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological review*.

[Nissen and Bullemer, 1987] Nissen, M. J. and Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive psychology*, 19(1):1–32.

[Ohara et al., 2001] Ohara, S., Mima, T., Baba, K., Ikeda, A., Kunieda, T., Matsumoto, R., Yamamoto, J., Matsuhashi, M., Nagamine, T., Hirasawa, K., et al. (2001). Increased synchronization of cortical oscillatory activities between human supplementary motor and primary sensorimotor areas during voluntary movements. *Journal of Neuroscience*, 21(23):9377–9386.

[O’Keefe, 1976] O’Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental Neurology*, 51(1):78–109.

[Oppenheimer and Kelso, 2015] Oppenheimer, D. M. and Kelso, E. (2015). Information processing as a paradigm for decision making. *Annual review of psychology*, 66:277–294.

[O’Reilly et al., 2008] O’Reilly, J. X., Mesulam, M. M., and Nobre, A. C. (2008). The cerebellum predicts the timing of perceptual events. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28(9):2252–60.

[Panksepp and Biven, 2012] Panksepp, J. and Biven, L. (2012). *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions (Norton Series on Interpersonal Neurobiology)*.

[Pasalar et al., 2006] Pasalar, S., Roitman, A. V., Durfee, W. K., and Ebner, T. J. (2006). Force field effects on cerebellar Purkinje cell discharge with implications for internal models. *Nature Neuroscience*, 9(11):1404–1411.

[Passingham et al., 2010] Passingham, R. E., Bengtsson, S. L., and Lau, H. C. (2010). Medial frontal cortex: from self-generated action to reflection on one’s own performance. *Trends in cognitive sciences*, 14(1):16–21.

[Paul Dean and John Porrill, 2010] Paul Dean, M. and John Porrill, M. (2010). The cerebellum as an adaptive filter: a general model? *Functional neurology*, 25(3):1–8.

[Pavlov, 1941] Pavlov, I. P. (1941). Lectures on conditioned reflexes. vol. ii. conditioned reflexes and psychiatry.

[Pavlov and Anrep, 2003] Pavlov, I. P. and Anrep, G. V. (2003). *Conditioned Reflexes*.

[Penfield and Jasper, 1954] Penfield, W. and Jasper, H. (1954). *Epilepsy and the functional anatomy of the human brain*.

[Pennartz et al., 2011] Pennartz, C. M. A., Ito, R., Verschure, P. F. M. J., Battaglia, F. P., and Robbins, T. W. (2011). The hippocampal-striatal axis in learning, prediction and goal-directed behavior. *Trends in neurosciences*, 34(10):548–59.

[Penny et al., 2008] Penny, W., Duzel, E., Miller, K., and Ojemann, J. (2008). Testing for nested oscillation. *Journal of neuroscience methods*, 174(1):50–61.

[Peterka, 2002] Peterka, R. J. (2002). Sensorimotor Integration in Human Postural Control. *J Neurophysiol*, 88(3):1097–1118.

[Pfaff, 1999] Pfaff, D. W. (1999). *Drive: Neurobiological and molecular mechanisms of sexual motivation. Cellular and molecular neuroscience.*

[Pfeifer et al., 2007] Pfeifer, R., Lungarella, M., and Iida, F. (2007). Self-Organization, Embodiment, and Biologically Inspired Robotics. *Science*, 318(5853):1088–1093.

[Phillips et al., 2014] Phillips, J. M., Vinck, M., Everling, S., and Womelsdorf, T. (2014). A long-range fronto-parietal 5- to 10-Hz network predicts ”top-down” controlled guidance in a task-switch paradigm. *Cerebral Cortex*, 24(8):1996–2008.

[Pravosudov and Smulders, 2010] Pravosudov, V. V. and Smulders, T. V. (2010). Integrating ecology, psychology and neurobiology within a food-hoarding paradigm. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1542):859–67.

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

[Prescott et al., 1999] Prescott, T. J., Redgrave, P., and Gurney, K. (1999). Layered Control Architectures in Robots and Vertebrates. *Adaptive Behavior*, 7(1):99–127.

[Ramchurn et al., 2014] Ramchurn, A., de Fockert, J. W., Mason, L., Darling, S., and Bunce, D. (2014). Intraindividual reaction time variability affects p300 amplitude rather than latency. *Frontiers in human neuroscience*, 8:557.

[Ramnani, 2006a] Ramnani, N. (2006a). The primate cortico-cerebellar system: anatomy and function. *Nature Reviews Neuroscience*, 7(7):511.

[Ramnani, 2006b] Ramnani, N. (2006b). The primate cortico-cerebellar system: Anatomy and function. *Nature Reviews Neuroscience*, 7(7):511–522.

[Ramnani, 2006c] Ramnani, N. (2006c). The primate cortico-cerebellar system: anatomy and function. *Nature reviews. Neuroscience*, 7(7):511–22.

[Ramnani, 2014] Ramnani, N. (2014). Automatic and controlled processing in the corticocerebellar system. In *Progress in brain research*, volume 210, pages 255–285. Elsevier.

[Ramnani and Passingham, 2001] Ramnani, N. and Passingham, R. (2001). Changes in the human brain during rhythm learning. *Journal of Cognitive Neuroscience*, 13(7):952–966.

[Ramón y Cajal, 1911] Ramón y Cajal, S. (1911). Histology of the nervous system, vol. ii.

[Ranganath and Blumenfeld, 2005] Ranganath, C. and Blumenfeld, R. S. (2005). Doubts about double dissociations between short-and long-term memory. *Trends in cognitive sciences*, 9(8):374–380.

[Reddi and Carpenter, 2000] Reddi, B. A. and Carpenter, R. H. (2000). The influence of urgency on decision time. *Nature neuroscience*, 3(8):827–30.

[Redish, 2016] Redish, A. D. (2016). Vicarious trial and error. *Nature Reviews Neuroscience*, 17(3):147–159.

[Rennó-Costa et al., 2010] Rennó-Costa, C., Lisman, J. E., and Verschure, P. F. M. J. (2010). The mechanism of rate remapping in the dentate gyrus. *Neuron*, 68(6):1051–8.

[Rescorla, 1988] Rescorla, R. A. (1988). Pavlovian conditioning: It’s not what you think it is. *American Psychologist*, 43(3):151.

[Rescorla et al., 1972] Rescorla, R. A., Wagner, A. R., et al. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, 2:64–99.

[Ridderinkhof et al., 2004] Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, 306(5695):443–447.

[Robertson, 2007] Robertson, E. M. (2007). The serial reaction time task: implicit motor skill learning? *Journal of Neuroscience*, 27(38):10073–10075.

[Roth et al., 2013a] Roth, M. J., Synofzik, M., and Lindner, A. (2013a). The cerebellum optimizes perceptual predictions about external sensory events. *Current Biology*, 23(10):930–935.

[Roth et al., 2013b] Roth, M. J., Synofzik, M., and Lindner, A. (2013b). The cerebellum optimizes perceptual predictions about external sensory events. *Current biology : CB*, 23(10):930–5.

[Ruan et al., 2007] Ruan, X., Ding, M., Gong, D., and Qiao, J. (2007). On-line adaptive control for inverted pendulum balancing based on feedback-error-learning. *Neurocomputing*, 70(4-6):770–776.

[Ruck et al., 2016] Ruck, M., Herreros, I., Maffei, G., Sánchez-Fibla, M., and Verschure, P. (2016). Learning to balance while reaching: A cerebellar-based control architecture for a self-balancing robot. In *Conference on Biomimetic and Biohybrid Systems*, pages 214–226. Springer.

[Rushworth and Hadland, 2002] Rushworth, M. F. S. and Hadland, K. A. (2002). Role of the Human Medial Frontal Cortex in Task Switching: A Combined fMRI and TMS Study. *Journal of . . .*, pages 2577–2592.

[Sainburg et al., 1999] Sainburg, R., Ghez, C., and Kalakanis, D. (1999). Intersegmental dynamics are controlled by sequential anticipatory, error correction, and postural mechanisms. *Journal of Neurophysiology*, 81(3):1045–1056.

[Sakai et al., 2004] Sakai, K., Hikosaka, O., and Nakamura, K. (2004). Emergence of rhythm during motor learning. *Trends in cognitive sciences*, 8(12):547–553.

[Sakai et al., 2002] Sakai, K., Ramnani, N., and Passingham, R. E. (2002). Learning of sequences of finger movements and timing: frontal lobe and action-oriented representation. *Journal of neurophysiology*, 88(4):2035–2046.

[Sánchez-Montañés et al., 2000] Sánchez-Montañés, M. A., Verschure, P. F. M. J., and König, P. (2000). Local and Global Gating of Synaptic Plasticity. *Neural Computation*, 12(3):519–529.

[Sandamirskaya et al., 2013] Sandamirskaya, Y., Zibner, S. K., Schneegans, S., and Schöner, G. (2013). Using Dynamic Field Theory to extend the embodiment stance toward higher cognition. *New Ideas in Psychology*, 31(3):322–339.

[Santos et al., 2010a] Santos, M. J., Kanekar, N., and Aruin, A. S. (2010a). The role of anticipatory postural adjustments in compensatory control of posture: 1. Electromyographic analysis. *Journal of electromyography and kinesiology : official journal of the International Society of Electrophysiological Kinesiology*, 20(3):388–97.

[Santos et al., 2010b] Santos, M. J., Kanekar, N., and Aruin, A. S. (2010b). The role of anticipatory postural adjustments in compensatory control of posture: 2. Biomechanical analysis. *Journal of electromyography and kinesiology : official journal of the International Society of Electrophysiological Kinesiology*, 20(3):398–405.

[Santos-Pata, Diogo Verschure, 2014] Santos-Pata, Diogo Verschure, P. F. (2014). *Hippocampal Based Model Reveals the Distinct Roles of Dentate Gyrus and CA3 during Robotic Spatial Navigation*, volume 8608 of *Lecture Notes in Computer Science*. Springer International Publishing, Cham.

[Savelsbergh et al., 1992] Savelsbergh, G., Whiting, H., Burden, A., and Bartlett, R. M. (1992). The role of predictive visual temporal information in the coordination of muscle activity in catching. *Experimental Brain Research*, 89(1):223–228.

[Schmahmann, 1996] Schmahmann, J. D. (1996). From movement to thought: Anatomic substrates of the cerebellar contribution to cognitive processing. *Human Brain Mapping*, 4(3):174–198.

[Schneider and Chein, 2003] Schneider, W. and Chein, J. M. (2003). Controlled & automatic processing: behavior, theory, and biological mechanisms. *Cognitive science*, 27(3):525–559.

[Schneider and Shiffrin, 1977] Schneider, W. and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychological review*, 84(1):1.

[Schultz, 2002] Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, 36(2):241–263.

[Serrien and Wiesendanger, 1999a] Serrien, D. J. and Wiesendanger, M. (1999a). Role of the cerebellum in tuning anticipatory and reactive grip force responses. *Journal of cognitive neuroscience*, 11(6):672–81.

[Serrien and Wiesendanger, 1999b] Serrien, D. J. and Wiesendanger, M. (1999b). Role of the cerebellum in tuning anticipatory and reactive grip force responses. *Journal of Cognitive Neuroscience*, 11(6):672–681.

[Shadmehr et al., 2016] Shadmehr, R., Huang, H. J., and Ahmed, A. A. (2016). A representation of effort in decision-making and motor control. *Current biology*, 26(14):1929–1934.

[Shadmehr and Mussa-Ivaldi, 1994] Shadmehr, R. and Mussa-Ivaldi, F. a. (1994). Adaptive representation of dynamics during learning of a motor task. *The Journal of Neuroscience*, 14(5):3208–3224.

[Shadmehr et al., 2010a] Shadmehr, R., Orban de Xivry, J. J., Xu-Wilson, M., and Shih, T.-Y. (2010a). Temporal discounting of reward and the cost of time in motor control. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(31):10507–16.

[Shadmehr et al., 2010b] Shadmehr, R., Smith, M. A., and Krakauer, J. W. (2010b). Error correction, sensory prediction, and adaptation in motor control. *Annual review of neuroscience*, 33:89–108.

[Shibata and Schaal, 2001] Shibata, T. and Schaal, S. (2001). Biomimetic gaze stabilization based on feedback-error-learning with nonparametric regression networks. *Neural Networks*, 14(2):201–216.

[Shidara et al., 1993] Shidara, M., Kawano, K., Gomi, H., and Kawato, M. (1993). Inverse-dynamics model eye movement control by Purkinje cells in the cerebellum. *Nature*, 365(6441):50–2.

[Shiffrin and Schneider, 1977] Shiffrin, R. M. and Schneider, W. (1977). Controlled and automatic human information processing: Ii. perceptual learning, automatic attending and a general theory. *Psychological review*, 84(2):127.

[Shiratori and Latash, 2001] Shiratori, T. and Latash, M. L. (2001). Anticipatory postural adjustments during load catching by standing subjects. *Clinical Neurophysiology*, 112(7):1250–1265.

[Siegel et al., 2009] Siegel, M., Warden, M. R., and Miller, E. K. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proceedings of the National Academy of Sciences*, 106(50):21341–21346.

[Simpson, 1955] Simpson, G. G. (1955). *Major features of evolution*. Columbia University Press: New York.

[Skinner, 1948] Skinner, B. F. (1948). ’Superstition’ in the pigeon. *Journal of Experimental Psychology,*.

[Sloman, 1996] Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119(1):3–22.

[Smith and Ratcliff, 2004] Smith, P. L. and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in neurosciences*, 27(3):161–8.

[Smith et al., 1990] Smith, R., Self, M., and Peter, C. (1990). Estimating Uncertain Spatial Relationships in Robotics. *Autonomous Robot Veichles*.

[Smulders et al., 2010] Smulders, T. V., Gould, K. L., and Leaver, L. A. (2010). Using ecology to guide the study of cognitive and neural mechanisms of different aspects of spatial memory in food-hoarding animals. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1542):883–900.

[Spear, 2014] Spear, N. E. (2014). *Information processing in animals: Memory mechanisms*. Psychology Press.

[Stapley et al., 2002] Stapley, P. J., Ting, L. H., Hulliger, M., and Macpherson, J. M. (2002). Automatic postural responses are delayed by pyridoxine-induced somatosensory loss. *Journal of Neuroscience*, 22(14):5803–5807.

[Stearns, 1985] Stearns, S. D. (1985). of aldapfive signal processing.

[Stout, 2010] Stout, D. (2010). The evolution of cognitive control. *Topics in Cognitive Science*, 2(4):614–630.

[Strick et al., 2009a] Strick, P. L., Dum, R. P., and Fiez, J. A. (2009a). Cerebellum and nonmotor function. *Annual review of neuroscience*, 32:413–434.

[Strick et al., 2009b] Strick, P. L., Dum, R. P., and Fiez, J. A. (2009b). Cerebellum and nonmotor function. *Annual review of neuroscience*, 32:413–34.

[Sutcliffe and de Lecea, 2002] Sutcliffe, J. G. and de Lecea, L. (2002). The hypocretins: setting the arousal threshold. *Nature reviews. Neuroscience*, 3(5):339–49.

[Sutton and Barto, 1998] Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge.

[Suvrathan et al., 2016] Suvrathan, A., Payne, H. L., and Raymond, J. L. (2016). Timing rules for synaptic plasticity matched to behavioral function. *Neuron*, 92(5):959–967.

[Swanson, 2012] Swanson, L. W. (2012). *Brain Architecture: Understanding the Basic Plan*.

[Sweeney-Reed et al., 2017] Sweeney-Reed, C. M., Zaehle, T., Voges, J., Schmitt, F. C., Buentjen, L., Borchardt, V., Walter, M., Hinrichs, H., Heinze, H.-J., Rugg, M. D., et al. (2017). Anterior thalamic high frequency band activity is coupled with theta oscillations at rest. *Frontiers in human neuroscience*, 11:358.

[Sweeney-Reed et al., 2014] Sweeney-Reed, C. M., Zaehle, T., Voges, J., Schmitt, F. C., Buentjen, L., Kopitzki, K., Esslinger, C., Hinrichs, H., Heinze, H.-J., Knight, R. T., et al. (2014). Corticothalamic phase synchrony and cross-frequency coupling predict human memory formation. *Elife*, 3.

[Szczepanski et al., 2014] Szczepanski, S. M., Crone, N. E., Kuperman, R. A., Auguste, K. I., Parvizi, J., and Knight, R. T. (2014). Dynamic changes in phase-amplitude coupling facilitate spatial attention control in fronto-parietal cortex. *PLoS biology*, 12(8):e1001936.

[Tanji, 1994] Tanji, J. (1994). The supplementary motor area in the cerebral cortex. *Neuroscience Research*, 19(3):251–268.

[Tenenbaum et al., 2011] Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285.

[Therrien and Bastian, 2015] Therrien, A. S. and Bastian, A. J. (2015). Cerebellar damage impairs internal predictions for sensory and motor function. *Current opinion in neurobiology*, 33:127–33.

[Thomson, 1982] Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096.

[Thorndike, 1927] Thorndike, E. L. (1927). The Law of Effect.

[Thoroughman and Shadmehr, 2000] Thoroughman, K. A. and Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. *Nature*, 407(6805):742–7.

[Timmann and Horak, 2001a] Timmann, D. and Horak, F. (2001a). Perturbed step initiation in cerebellar subjects: 2. modification of anticipatory postural adjustments. *Experimental brain research*, 141(1):110–120.

[Timmann and Horak, 2001b] Timmann, D. and Horak, F. B. (2001b). Perturbed step initiation in cerebellar subjects: 2. Modification of anticipatory postural adjustments. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, 141(1):110–20.

[Tinbergen,] Tinbergen, N. *The study of instinct.*

[Todorov and Jordan, 2002] Todorov, E. and Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature neuroscience*, 5(11):1226.

[Tolman,] Tolman, E. C. Cognitive maps in rats and men.

[Tolman, 1938] Tolman, E. C. (1938). a t a Choice Point. 5(2):267–272.

[Tolman, 1951] Tolman, E. C. (1951). *Purposive behavior in animals and men.*

[Tolman and Honzik, 1930] Tolman, E. C. and Honzik, C. H. (1930). Introduction and removal of reward, and maze performance in rats. *University of California publications in psychology.*

[Tseng et al., 2007] Tseng, Y.-w., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., and Bastian, A. J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *Journal of neurophysiology*, 98(1):54–62.

[Tversky and Kahneman, 1974] Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131.

[Ulrich, 2002] Ulrich, D. (2002). Dendritic resonance in rat neocortical pyramidal cells. *Journal of neurophysiology*, 87(6):2753–2759.

[Uusisaari and de Schutter, 2011] Uusisaari, M. and de Schutter, E. (2011). The mysterious microcircuitry of the cerebellar nuclei. *Journal of Physiology*, 589(14):3441–3457.

[van der Meer et al., 2012] van der Meer, M., Kurth-Nelson, Z., and Redish, A. D. (2012). Information Processing in Decision-Making Systems. *The Neuroscientist*, 18(4):342–359.

[Van Der Smagt, 1998] Van Der Smagt, P. (1998). Cerebellar Control of Robot Arms. *Connection Science*, 10(3-4):301–320.

[van Dorp and De Zeeuw, 2014] van Dorp, S. and De Zeeuw, C. I. (2014). Variable timing of synaptic transmission in cerebellar unipolar brush cells. *Proceedings of the National Academy of Sciences of the United States of America*, pages 1314219111–.

[van Hemmen et al., 2014] van Hemmen, J. L., Schüz, A., and Aertsen, A. (2014). Structural aspects of biological cybernetics: Valentino braitenberg, neuroanatomy, and brain function. *Biological cybernetics*, 108(5):517–525.

[Vansteenkiste et al., 2014] Vansteenkiste, P., Van Hamme, D., Veelaert, P., Philippaerts, R., Cardon, G., and Lenoir, M. (2014). Cycling around a curve: the effect of cycling speed on steering and gaze behavior. *PloS one*, 9(7):e102792.

[Varela et al., 2017] Varela, F. J., Thompson, E., and Rosch, E. (2017). *The embodied mind: Cognitive science and human experience*. MIT press.

[Verschure, 2012] Verschure, P. F. (2012). Distributed Adaptive Control: A theory of the Mind, Brain, Body Nexus. *Biologically Inspired Cognitive Architectures*, 1:55–72.

[Verschure et al., 2014a] Verschure, P. F., Pennartz, C. M., and Pezzulo, G. (2014a). The why, what, where, when and how of goal-directed choice: neuronal and computational principles. *Phil. Trans. R. Soc. B*, 369(1655):20130483.

[Verschure and Voegtlin, 1998] Verschure, P. F. and Voegtlin, T. (1998). A bottom up approach towards the acquisition and expression of sequential representations applied to a behaving real-world device: Distributed Adaptive Control III. *Neural Networks*, 11(7-8):1531–1549.

[Verschure et al., 2003a] Verschure, P. F., Voegtlin, T., and Douglas, R. J. (2003a). Environmentally mediated synergy between perception and behaviour in mobile robots. *Nature*, 425(6958):620.

[Verschure et al., 1995] Verschure, P. F., Wray, J., Sporns, O., Tononi, G., and Edelman, G. M. (1995). Multilevel analysis of classical conditioning in a behaving real world artifact. *Robotics and Autonomous Systems*, 16(2-4):247–265.

[Verschure et al., 2014b] Verschure, P. F. M. J., Pennartz, C. M. A., and Pezzulo, G. (2014b). The why, what, where, when and how of goal-directed choice: neuronal and computational principles. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1655):20130483.

[Verschure et al., 2003b] Verschure, P. F. M. J., Voegtlin, T., and Douglas, R. J. (2003b). Environmentally mediated synergy between perception and behaviour in mobile robots. *Nature*, 425(6958):620–4.

[Volkow et al., 2011] Volkow, N. D., Wang, G.-J., and Baler, R. D. (2011). Reward, dopamine and the control of food intake: implications for obesity. *Trends in cognitive sciences*, 15(1):37–46.

[Voloh et al., 2015] Voloh, B., Valiante, T. A., Everling, S., and Womelsdorf, T. (2015). Theta–gamma coordination between anterior cingulate and prefrontal cortex indexes correct attention shifts. *Proceedings of the National Academy of Sciences*, 112(27):8457–8462.

[Voytek et al., 2010] Voytek, B., Canolty, R. T., Shestyuk, A., Crone, N., Parvizi, J., and Knight, R. T. (2010). Shifts in gamma phase–amplitude coupling frequency from theta to alpha over posterior cortex during visual tasks. *Frontiers in human neuroscience*, 4:191.

[Voytek et al., 2015] Voytek, B., Kayser, A. S., Badre, D., Fegen, D., Chang, E. F., Crone, N. E., Parvizi, J., Knight, R. T., and D’esposito, M. (2015). Oscillatory dynamics coordinating human frontal networks in support of goal maintenance. *Nature neuroscience*, 18(9):1318.

[Walter, 1963] Walter, W. G. (1963). The development and significance of cybernetics’. *Anarch. 25. A J. Anarch. Ideas*, 3(3):75–89.

[Waniek et al., 2013] Waniek, N., Denk, C., and Conradt, J. (2013). GRIDMAP - From brains to technical implementations.

[Warwick et al., 1973] Warwick, R., Williams, P. L., and Gray, H. (1973). *Gray’s anatomy*. Longman.

[Weinberger, 2004] Weinberger, N. M. (2004). Specific long-term memory traces in primary auditory cortex. *Nature reviews. Neuroscience*, 5(4):279–90.

[Widrow et al., 1967] Widrow, B., Mantey, P., Griffiths, L., and Goode, B. (1967). Adaptive antenna systems. *Proceedings of the IEEE*, 55(12):2143–2159.

[Wiener, 1948] Wiener, N. (1948). Cybernetics. *Scientific American*, 179(5):14–19.

[Wiener, 1961] Wiener, N. (1961). *Cybernetics or Control and Communication in the Animal and the Machine*, volume 25. MIT press.

[Wiestler et al., 2011] Wiestler, T., McGonigle, D. J., and Diedrichsen, J. (2011). Integration of sensory and motor representations of single fingers in the human cerebellum. *Journal of neurophysiology*, 105(6):3042–53.

[Witney et al., 1999] Witney, A. G., Goodbody, S. J., and Wolpert, D. M. (1999). Predictive motor learning of temporal delays. *Journal of Neurophysiology*, 82(5):2039–2048.

[Wolpert et al., 2011] Wolpert, D. M., Diedrichsen, J., and Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature reviews. Neuroscience*, 12(12):739–51.

[Wolpert and Kawato, 1998] Wolpert, D. M. and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural networks*, 11(7-8):1317–1329.

[Wolpert et al., 1998a] Wolpert, D. M., Miall, R., and Kawato, M. (1998a). Internal models in the cerebellum. *Trends in Cognitive Sciences*, 2(9):338–347.

[Wolpert et al., 1998b] Wolpert, D. M., Miall, R. C., and Kawato, M. (1998b). Internal models in the cerebellum. *Trends in cognitive sciences*, 2(9):338–347.

[Wong and Wang, 2006] Wong, K.-F. and Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4):1314–1328.

[Woods, 1991] Woods, S. C. (1991). The eating paradox: how we tolerate food. *Psychological review*, 98(4):488–505.

[Woods and Ramsay, 2000] Woods, S. C. and Ramsay, D. S. (2000). Pavlovian influences over food and drug intake. *Behavioural brain research*, 110(1-2):175–82.

[Xu et al., 2004] Xu, D., Carlton, L. G., and Rosengren, K. S. (2004). Anticipatory postural adjustments for altering direction during walking. *Journal of motor behavior*, 36(3):316–326.

[Yoon et al., 2013] Yoon, K., Buice, M. A., Barry, C., Hayman, R., Burgess, N., and Fiete, I. R. (2013). Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature neuroscience*, 16(8):1077–84.

[Zamble et al., 1985] Zamble, E., Hadad, G. M., Mitchell, J. B., and Cutmore, T. R. (1985). Pavlovian conditioning of sexual arousal: first- and second-order effects. *Journal of experimental psychology. Animal behavior processes*, 11(4):598–610.

[Zavala et al., 2016] Zavala, B., Tan, H., Ashkan, K., Foltynie, T., Limousin, P., Zrinzo, L., Zaghloul, K., and Brown, P. (2016). Human subthalamic nucleus–medial frontal cortex theta phase coherence is involved in conflict and error related cortical monitoring. *NeuroImage*, 137:178–187.